GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 16, 2003, 03:49:46 ; Search time 1820.86 Seconds
(without alignments)
15998.962 Million cell updates/sec

Title: US-09-981-353-24
Perfect score: 1001
Sequence: 1 atgattacgaattcgagctc.....gtgagcaaaaggccagcaaa 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	582	58.1	632	9	BC029149	BC029149 Homo sapi
2	545.8	54.5	560	6	AR182207	AR182207 Sequence
3	545.8	54.5	560	6	I95747	I95747 Sequence 10
4	397.6	39.7	105485	9	HUAC002301	AC002301 Homo sapi
c 5	387.4	38.7	174007	9	AC023831	AC023831 Homo sapi
6	387.4	38.7	175599	2	AC009133	AC009133 Homo sapi
7	368.8	36.8	673	10	BC031800	BC031800 Mus muscu
c 8	357.2	35.7	3409	12	CVU94951	U94951 Cloning vec
c 9	357.2	35.7	3595	12	AF174493	AF174493 Cloning v
c 10	357.2	35.7	3733	12	AF146395	AF146395 Cloning v
c 11	357.2	35.7	4630	12	CVU35316	U35316 Cloning vec
c 12	357.2	35.7	4630	12	CVU94950	U94950 Cloning vec
c 13	355.8	35.5	5086	6	AX107161	AX107161 Sequence
c 14	355.8	35.5	5086	12	AF438204	AF438204 Positive
15	352.6	35.2	3417	12	AF153422	AF153422 Cloning v
16	352.6	35.2	9152	12	AF527485	AF527485 Chloropla
c 17	352.4	35.2	5925	6	AX009712	AX009712 Sequence
c 18	352.4	35.2	5925	6	AX010972	AX010972 Sequence
c 19	352.4	35.2	6680	6	AX287967	AX287967 Sequence
20	352.2	35.2	3789	6	AR130382	AR130382 Sequence
21	352	35.2	3950	12	EVU81273	U81273 Expression
22	352	35.2	3966	12	EVU81274	U81274 Expression
23	352	35.2	3970	12	EVU81275	U81275 Expression
24	352	35.2	3983	6	AX195438	AX195438 Sequence
c 25	351.8	35.1	8710	6	I60241	I60241 Sequence 3
c 26	351.8	35.1	9019	6	I60242	I60242 Sequence 4
c 27	351.6	35.1	4190	12	XXU13848	U13848 pExCell clo
c 28	351.4	35.1	4591	12	SYNPEZZ18A	M74186 Cloning vec
29	351.4	35.1	8062	6	AX114872	AX114872 Sequence
30	351.4	35.1	8153	6	AX114871	AX114871 Sequence
31	351.2	35.1	2939	12	SYNPLINKY	D50400 Phagemid pL
32	351.2	35.1	3837	12	SYNPPDSZ	D50401 Phagemid pP
c 33	351	35.1	8800	6	AX469663	AX469663 Sequence
c 34	351	35.1	9208	12	AF064067	AF064067 Expressio
c 35	351	35.1	9211	12	AF064066	AF064066 Expressio
c 36	351	35.1	9215	12	AF049063	AF049063 Expressio
c 37	351	35.1	9956	12	AF049064	AF049064 Expressio
c 38	351	35.1	10042	6	AX469664	AX469664 Sequence
39	350.8	35.0	584	6	AX063012	AX063012 Sequence
40	350.8	35.0	584	6	AX367929	AX367929 Sequence
41	350.8	35.0	655	6	AX351379	AX351379 Sequence
42	350.8	35.0	738	6	AR036905	AR036905 Sequence
43	350.8	35.0	738	6	AR141144	AR141144 Sequence
44	350.8	35.0	738	6	AR181919	AR181919 Sequence
45	350.8	35.0	742	6	AX197652	AX197652 Sequence

ALIGNMENTS

RESULT 1
BC029149
LOCUS BC029149 632 bp mRNA linear PRI 16-MAY-2002
DEFINITION Homo sapiens, similar to Homolog of rat Zymogen granule membrane protein, clone MGC:34820 IMAGE:5187790, mRNA, complete cds.
ACCESSION BC029149
VERSION BC029149.1 GI:20810120
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 632)
AUTHORS Strausberg,R.
TITLE Direct Submission

```
  JOURNAL   Submitted (01-MAY-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amg@bcm.tmc.edu
            Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
            Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
            Richards, S., Gibbs, R.A.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 50 Row: h Column: 15
            This clone was selected for full length sequencing because it
            passed the following selection criteria: GenomeScan gene
            prediction.
FEATURES             Location/Qualifiers
     source          1. .632
                     /organism="Homo sapiens"
                     /db_xref="LocusID:123887"
                     /db_xref="taxon:9606"
                     /clone="MGC:34820 IMAGE:5187790"
                     /tissue_type="Colon, Kidney, Stomach, adult, whole pooled"
                     /clone_lib="NIH_MGC_116"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS             7. .510
                     /codon_start=1
                     /product="similar to Homolog of rat Zymogen granule
                     membrane protein"
                     /protein_id="AAH29149.1"
                     /db_xref="GI:20810121"
                     /translation="MLTVALLALLCASASGNAIQARSSSYSGEYGSGGGKRFSHSGNQ
                     LDGPITALRVRVNTYYIVGLQVRYGKVWSDYVGGRNGDLEEIFLHPGESVIQVSGKYK
                     WYLKKLVFVTDKGRYLSFGKDSGTSFNAVPLHPNTVLRFISGRSGSLIDAIGLHWDVY
                     PTSCSRC"
BASE COUNT      157 a    165 c    161 g    149 t
ORIGIN

  Query Match           58.1%;  Score 582;  DB 9;  Length 632;
  Best Local Similarity  100.0%;  Pred. No. 4e-159;
  Matches  582;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         28 CCCAGAATGTTGACAGTCGCTCTCCTAGCCCTTCTCTGTGCCTCAGCCTCTGGCAATGCC 87
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCCAGAATGTTGACAGTCGCTCTCCTAGCCCTTCTCTGTGCCTCAGCCTCTGGCAATGCC 60

Qy         88 ATTCAGGCCAGGTCTTCCTCCTATAGTGGAGAGTATGGAAGTGGTGGTGGAAAGCGATTC 147
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ATTCAGGCCAGGTCTTCCTCCTATAGTGGAGAGTATGGAAGTGGTGGTGGAAAGCGATTC 120

Qy        148 TCTCATTCTGGCAACCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACACA 207
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCTCATTCTGGCAACCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACACA 180

Qy        208 TACTACATCGTAGGTCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGGT 267
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TACTACATCGTAGGTCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGGT 240

Qy        268 CGCAACGGAGACCTGGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTCT 327
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CGCAACGGAGACCTGGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTCT 300

Qy        328 GGGAAGTACAAGTGGTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCTG 387
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGGAAGTACAAGTGGTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCTG 360

Qy        388 TCTTTTGGGAAAGACAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGTG 447
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCTTTTGGGAAAGACAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGTG 420

Qy        448 CTCCGCTTCATCAGTGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGAT 507
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTCCGCTTCATCAGTGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGAT 480

Qy        508 GTTTACCCCACTAGCTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGAT 567
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GTTTACCCCACTAGCTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGAT 540

Qy        568 GAGGAGTAAGAACTCCCTTATCACTAACCCCCATCCAAATGG 609
              ||||||||||||||||||||||||||||||||||||||||||
Db        541 GAGGAGTAAGAACTCCCTTATCACTAACCCCCATCCAAATGG 582


RESULT 2
AR182207
LOCUS       AR182207                 560 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 10 from patent US 6337195.
ACCESSION   AR182207
VERSION     AR182207.1  GI:20225123
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 560)
  AUTHORS   Yu,G.-L. and Rosen,C.
  TITLE     Colon specific genes and proteins
  JOURNAL   Patent: US 6337195-A 10 08-JAN-2002;
FEATURES             Location/Qualifiers
     source          1. .560
                     /organism="unknown"
BASE COUNT      113 a    157 c    152 g    138 t
ORIGIN

  Query Match           54.5%;  Score 545.8;  DB 6;  Length 560;
  Best Local Similarity  99.5%;  Pred. No. 1.6e-148;
  Matches  558;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy         42 AGTCGCTCTCCTAGCCCTTCTCTGTGCCTCAGCCTCTGGCAATGCCATTCAGGCCAGGTC 101
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db          1 AGTCGCTCTCCTAGCCCTTCTCTGTGCCTCACCCTCTGGCAATGCCATTCAGGCCAGGTC 60

Qy        102 TTCCTCCTATAGTGGAGAGTATGGAAGTGGTGGTGGAAAGCGATTCTCTCATTCTGGCAA 161
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTCCTCCTATAGTGGAGAGTATGGAGGTGGTGGTGGAAAGCGATTCTCTCATTCTGGCAA 120

Qy        162 CCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACACATACTACATCGTAGG 221
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACACATACTACATCGTAGG 180

Qy        222 TCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGGTCGCAACGGAGACCT 281
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGGTCGCAACGGAGACCT 240

Qy        282 GGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTCTGGGAAGTACAAGTG 341
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTCTGGGAAGTACAAGTG 300

Qy        342 GTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCTGTCTTTTGGGAAAGA 401
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCTGTCTTTTGGGAAAGA 360

Qy        402 CAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGTGCTCCGCTTCATCAG 461
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGTGCTCCGCTTCATCAG 420

Qy        462 TGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGATGTTTACCCCACTAG 521
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGATGTTTACCCCACTAG 480

Qy        522 CTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGATGAGGAGTAAGAACT 581
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGATGAGGAGTAAGAACT 540

Qy        582 CCCTTATCACTAACCCCCATC 602
                 ||||||||||||||||||||
Db        541 -CCTTATCACTAACCCCCATC 560


RESULT 3
I95747
LOCUS       I95747                   560 bp    DNA     linear   PAT 01-DEC-1998
DEFINITION  Sequence 10 from patent US 5733748.
ACCESSION   I95747
VERSION     I95747.1  GI:3940217
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 560)
  AUTHORS   Yu,G.-L. and Rosen,C.
  TITLE     Colon specific genes and proteins
  JOURNAL   Patent: US 5733748-A 10 31-MAR-1998;
FEATURES             Location/Qualifiers
     source          1. .560
                     /organism="unknown"
BASE COUNT      113 a    157 c    152 g    138 t
ORIGIN

  Query Match           54.5%;  Score 545.8;  DB 6;  Length 560;
  Best Local Similarity  99.5%;  Pred. No. 1.6e-148;
  Matches  558;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy         42 AGTCGCTCTCCTAGCCCTTCTCTGTGCCTCAGCCTCTGGCAATGCCATTCAGGCCAGGTC 101
               ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db          1 AGTCGCTCTCCTAGCCCTTCTCTGTGCCTCACCCTCTGGCAATGCCATTCAGGCCAGGTC 60

Qy        102 TTCCTCCTATAGTGGAGAGTATGGAAGTGGTGGTGGAAAGCGATTCTCTCATTCTGGCAA 161
               ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db         61 TTCCTCCTATAGTGGAGAGTATGGAGGTGGTGGTGGAAAGCGATTCTCTCATTCTGGCAA 120

Qy        162 CCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACACATACTACATCGTAGG 221
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACACATACTACATCGTAGG 180

Qy        222 TCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGGTCGCAACGGAGACCT 281
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGGTCGCAACGGAGACCT 240

Qy        282 GGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTCTGGGAAGTACAAGTG 341
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTCTGGGAAGTACAAGTG 300

Qy        342 GTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCTGTCTTTTGGGAAAGA 401
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCTGTCTTTTGGGAAAGA 360

Qy        402 CAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGTGCTCCGCTTCATCAG 461
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGTGCTCCGCTTCATCAG 420

Qy        462 TGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGATGTTTACCCCACTAG 521
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGATGTTTACCCCACTAG 480

Qy        522 CTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGATGAGGAGTAAGAACT 581
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGATGAGGAGTAAGAACT 540

Qy        582 CCCTTATCACTAACCCCCATC 602
                 ||||||||||||||||||||
Db        541 -CCTTATCACTAACCCCCATC 560


RESULT 4
HUAC002301
LOCUS       HUAC002301             105485 bp    DNA     linear   PRI 23-NOV-1999
DEFINITION  Homo sapiens Chromosome 16 BAC clone CIT987SK-A-328A3, complete
            sequence.
ACCESSION   AC002301
VERSION     AC002301.1  GI:2979567
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 105485)
  AUTHORS   Loftus,B.J., Kim,U.J., Sneddon,V.P., Kalush,F., Brandon,R.,
            Fuhrmann,J., Mason,T., Crosby,M.L., Barnstead,M., Cronin,L.,
            Deslattes Mays,A., Cao,Y., Xu,R.X., Kang,H.L., Mitchell,S.,
            Eichler,E.E., Harris,P.C., Venter,J.C. and Adams,M.D.
  TITLE     Genome duplications and other features in 12 Mb of DNA sequence
            from human chromosome 16p and 16q
  JOURNAL   Genomics 60 (3), 295-308 (1999)
  MEDLINE   99425270
   PUBMED   10493829
REFERENCE   2  (bases 1 to 105485)
  AUTHORS   Adams,M.D., Loftus,B.J., Zhou,L., Crosby,M., Fuhrmann,J.,
            Mason,T.M., Brandon,R., Kim,U.J., Kerlavage,A.R. and Venter,J.C.
  TITLE     Homo sapiens Chromosome 16 BAC clone CIT987SK-A-328A3
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 105485)
  AUTHORS   Adams,M.D. and Loftus,B.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUN-1997) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA, Email:
            mdadams@tigr.org
REFERENCE   4  (bases 1 to 105485)
  AUTHORS   Adams,M.D. and Loftus,B.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAR-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
REFERENCE   5  (bases 1 to 105485)
  AUTHORS   Adams,M.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAR-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT     On Mar 20, 1998 this sequence version replaced gi:2951943.
            Address all correspondence to: Mark Adams The Institute for Genomic
            Research 9712 Medical Center Dr, Rockville, MD 20850, USA e-mail
            address: mdadams@tigr.org.   The orientation of the sequence is
            from SP6 end to T7 end. Genes were identified by a combination of
            five methods including: XGRAIL (available by anonymous ftp from
            arthur.epm.ornl.gov), Genefinder (Phil Green, University of
            Washington), Genscan  (Chris Burge,
            http://gnomic.stanford.Edu/~chris/GENSCANW.html)searches of the
            complete sequence against a peptide database,and the Human gene
            Index database at TIGR (http://www.tigr.org/tdb/hgi/hgi.html).
            Genes without pepetide homolgy having spliced EST hits are termed
            'Unknown gene product'. Genes encoding tRNAs are predicted by
            tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
FEATURES             Location/Qualifiers
     source          1. .105485
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="+16p11.2"
                     /clone="A-328A3"
     STS             3032. .3149
                     /note="7608, STS1-cSRL-24g1-uA/cSRL-24g1-uZ, Chr. -, Homo
                     sapiens"
                     /db_xref="dbSTS:G02122"
     STS             14885. .15002
```

```
                /note="7608, STS1-cSRL-24g1-uA/cSRL-24g1-uZ, Chr. -, Homo
                sapiens"
                /db_xref="dbSTS:G02122"
     STS        75574. .75694
                /note="22506, SHGC-5801, Chr. 17, Homo sapiens"
                /db_xref="dbSTS:G14087"
     gene       <77884. .79132
                /gene="A-328A3.1"
     mRNA       join(<77884. .77938,78122. .78254,78721. .79132)
                /gene="A-328A3.1"
     CDS        join(77884. .77938,78122. .78254,78721. .79036)
                /gene="A-328A3.1"
                /codon_start=1
                /product="Homolog of rat Zymogen granule membrane protein"
                /protein_id="AAC08708.1"
                /db_xref="GI:2979568"
                /translation="MLTVALLALLCASASGNAIQARSSSYSGEYGSGGGKRFSHSGNQ
                LDGPITALRVRVNTYYIVGLQVRYGKVWSDYVGGRNGDLEEIFLHPGESVIQVSGKYK
                WYLKKLVFVTDKGRYLSFGKDSGTSFNAVPLHPNTVLRFISGRSGSLIDAIGLHWDVY
                PTSCSRC"
     STS        87451. .87568
                /note="7608, STS1-cSRL-24g1-uA/cSRL-24g1-uZ, Chr. -, Homo
                sapiens"
                /db_xref="dbSTS:G02122"
     gene       89374. .>97443
                /gene="A-328A3.2"
                /note="(partial)"
     mRNA       join(89374. .89467,95545. .95740,97022. .97149,97241. .>97443)
                /gene="A-328A3.2"
     CDS        join(89398. .89467,95545. .95740,97022. .97149,97241. .97443)
                /gene="A-328A3.2"
                /codon_start=1
                /product="Similar to kinesin-like DNA binding protein
                (KID)"
                /protein_id="AAC08709.1"
                /db_xref="GI:2979569"
                /translation="MAAGGSTQQRRREMAAASAAAISGAGRCRLSKIGATRRPPPARV
                RVAVRLRPFVDGTAGASDPPCVRGMDSCSLEIANWRNHQETLKYQFDAFYGERSTQQD
                IYAGSVQPILRHLLEGQNASVLAYGPTGAGKTHTMLGSPEQPGVIPRALMDLLQLTRE
                EGAEGRPWALSVTMSYLEIYQEKVRPRAGWERSNKGSK"
     gene       102130. .104031
                /gene="A-328A3.5"
     mRNA       join(102130. .102255,102645. .102712,103238. .104031)
                /gene="A-328A3.5"
     mRNA       join(102130. .102255,102645. .102712,103238. .103673,
                103910. .104031)
                /gene="A-328A3.5"
                /note="102132"
     mRNA       join(102130. .102255,102645. .102712,103461. .103673,
                103762. .103821,103910. .104031)
                /gene="A-328A3.5"
     CDS        join(102132. .102255,102645. .102712,103461. .103673,
                103762. .103821,103910. .103957)
                /gene="A-328A3.5"
                /codon_start=1
                /product="oriP binding protein like gene product"
                /protein_id="AAC08711.1"
                /db_xref="GI:2979571"
                /translation="MLAQKAEEKENHCPTMLRPLSHRTVTRAKPLKKAVVMPLQLIQE
                QAASPNAEIHILKNKGRKRKLESLDALEPEEKAEDCWELQISPELLAHGRQKILDLLN
                EGSARDLRSLQRIGPKKAQLIVGWRELHGPFSQVEDLERVEGITGKQMESFLKANILG
                LAAGQRCGAS"
     CDS        join(102132. .102255,102645. .102712,103238. .103264)
                /gene="A-328A3.5"
                /codon_start=1
                /product="oriP binding protein like gene product"
                /protein_id="AAC08712.1"
                /db_xref="GI:2979572"
                /translation="MLAQKAEEKENHCPTMLRPLSHRTVTRAKPLKKAVVMPLQLIQE
                QAASPNAEIHILKNKGRKRKGANITTGY"
     misc_feature 104673. .104723
                /note="50 N's represent a gap of approximately 800 bp"
     gene       <105069. .>105276
                /gene="A-328A3.4"
                /note="(partial)"
     mRNA       <105069. .>105276
                /gene="A-328A3.4"
     CDS        105078. .>105276
                /gene="A-328A3.4"
                /codon_start=1
                /product="Similar to fragment of MAZ DNA binding protein"
                /protein_id="AAC08710.1"
                /db_xref="GI:2979570"
                /translation="MCGKAFRDVYHLNRHKLSHSDEKPYQCPVCQQRFKRKDRMSYHV
                RSHDGAVHKPYNCSHCGKSFSR"
BASE COUNT    27103 a  25382 c  25633 g  27316 t     51 others
ORIGIN

  Query Match           39.7%;  Score 397.6;  DB 9;  Length 105485;
  Best Local Similarity   88.5%;  Pred. No. 7.9e-105;
  Matches  406;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy        213 CATCGTAGGTCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGGTCGCAA 272
              | || |  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78712 CTTCCTCAGTCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGGTCGCAA 78771

Qy        273 CGGAGACCTGGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTCTGGGAA 332
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78772 CGGAGACCTGGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTCTGGGAA 78831

Qy        333 GTACAAGTGGTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCTGTCTTT 392
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78832 GTACAAGTGGTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCTGTCTTT 78891

Qy        393 TGGGAAAGACAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGTGCTCCG 452
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78892 TGGGAAAGACAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGTGCTCCG 78951

Qy        453 CTTCATCAGTGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGATGTTTA 512
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78952 CTTCATCAGTGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGATGTTTA 79011

Qy        513 CCCCACTAGCTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGATGAGGA 572
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79012 CCCCACTAGCTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGATGAGGA 79071

Qy        573 GTAAGAACTCCCTTATCACTAACCCCCATCCAAATGGNNNNNNNNNNNNNNNNNNNNNNN 632
              ||||||||||||||||||||||||||||||||||
Db      79072 GTAAGAACTCCCTTATCACTAACCCCCATCCAAATGGCTCAATAAAAAAATATGGTTAAG 79131

Qy        633 NNNNNNNNNNNNNNNNTCCTGTGTGAAATTGTTATCCGC 671
                              | | ||||     | |||| ||
Db      79132 GCTAGTCTGTGTGGGGGCATCTGTGGCTGGGATATCTGC 79170


RESULT 5
AC023831/c
LOCUS       AC023831              174007 bp    DNA     linear   PRI 23-AUG-2002
DEFINITION  Homo sapiens chromosome 16 clone RP11-74E23, complete sequence.
ACCESSION   AC023831
VERSION     AC023831.9  GI:22450570
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 174007)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 174007)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-FEB-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 174007)
```

```
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUN-2002) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   4  (bases 1 to 174007)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-AUG-2002) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Aug 23, 2002 this sequence version replaced gi:21397291.
            Sequence Quality Assessment:
            This entry has been annotated with sequence quality
            estimates computed by the Phrap assembly program.
            All manually edited bases have been reduced to quality zero.
            Quality levels above 40 are expected to have less than
            1 error in 10,000 bp.
            Base-by-base quality values are not generally visible from the
            GenBank flat file format but are available as part
            of this entry's ASN.1 file.
            ----------------------.
FEATURES             Location/Qualifiers
     source          1. .174007
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-74E23"
     misc_feature    1. .174007
                     /note="clone overlap not submitted"
BASE COUNT    44719 a  42779 c  42761 g  43748 t
ORIGIN

  Query Match            38.7%;  Score 387.4;  DB 9;  Length 174007;
  Best Local Similarity  98.5%;  Pred. No. 8.2e-102;
  Matches  391;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy        213 CATCGTAGGTCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGGTCGCAA 272
              | || |  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26047 CTTCCTCAGTCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGGTCGCAA 25988

Qy        273 CGGAGACCTGGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTCTGGGAA 332
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25987 CGGAGACCTGGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTCTGGGAA 25928

Qy        333 GTACAAGTGGTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCTGTCTTT 392
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db      25927 GTACAAGTGGTACCTGAAGAAGCTGTTATTTGTGACAGACAAGGGCCGCTATCTGTCTTT 25868

Qy        393 TGGGAAAGACAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGTGCTCCG 452
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25867 TGGGAAAGACAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGTGCTCCG 25808

Qy        453 CTTCATCAGTGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGATGTTTA 512
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25807 CTTCATCAGTGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGATGTTTA 25748

Qy        513 CCCCACTAGCTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGATGAGGA 572
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25747 CCCCAGTAGCTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGATGAGGA 25688

Qy        573 GTAAGAACTCCCTTATCACTAACCCCCATCCAAATGG 609
              |||||||||||||||||||||||||||||||||||||
Db      25687 GTAAGAACTCCCTTATCACTAACCCCCATCCAAATGG 25651


RESULT 6
AC009133
LOCUS       AC009133               175599 bp    DNA     linear   HTG 25-APR-2001
DEFINITION  Homo sapiens chromosome 16 clone RP11-504I2, WORKING DRAFT
            SEQUENCE, 5 ordered pieces.
ACCESSION   AC009133
VERSION     AC009133.6  GI:13786313
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 175599)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 175599)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Apr 25, 2001 this sequence version replaced gi:7689951.
            -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 600095
            Center clone name: RPCI-11_504I2
            -------
            Summary Statistics
            Consensus quality: 172472 bases at least Q40
            Consensus quality: 174610 bases at least Q30
            Consensus quality: 175032 bases at least Q20
            Estimated insert size: 190110; agarose-fp estimation
            Estimated insert size: 175199; sum-of-contigs estimation
            Quality coverage:  10.66 in Q20 bases; agarose-fp estimation
            Quality coverage:  11.57 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1    22409: contig of 22409 bp in length
            *    22410    22509: gap of unknown length
            *    22510    31954: contig of 9445 bp in length
            *    31955    32054: gap of unknown length
            *    32055   116297: contig of 84243 bp in length
            *   116298   116397: gap of unknown length
            *   116398   145562: contig of 29165 bp in length
            *   145563   145662: gap of unknown length
            *   145663   175599: contig of 29937 bp in length.
FEATURES             Location/Qualifiers
     source          1. .175599
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-504I2"
                     /clone_lib="RPCI human BAC library 11"
BASE COUNT    43184 a  44157 c  44428 g  43430 t    400 others
ORIGIN

  Query Match            38.7%;  Score 387.4;  DB 2;  Length 175599;
  Best Local Similarity  98.5%;  Pred. No. 8.2e-102;
  Matches  391;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy        213 CATCGTAGGTCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGGTCGCAA 272
              | || |  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      90251 CTTCCTCAGTCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGGTCGCAA 90310

Qy        273 CGGAGACCTGGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTCTGGGAA 332
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      90311 CGGAGACCTGGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTCTGGGAA 90370

Qy        333 GTACAAGTGGTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCTGTCTTT 392
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db      90371 GTACAAGTGGTACCTGAAGAAGCTGTTATTTGTGACAGACAAGGGCCGCTATCTGTCTTT 90430
```

```
Qy        393 TGGGAAAGACAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGTGCTCCG 452
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      90431 TGGGAAAGACAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGTGCTCCG 90490

Qy        453 CTTCATCAGTGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGATGTTTA 512
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      90491 CTTCATCAGTGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGATGTTTA 90550

Qy        513 CCCCACTAGCTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGATGAGGA 572
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      90551 CCCCAGTAGCTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGATGAGGA 90610

Qy        573 GTAAGAACTCCCTTATCACTAACCCCCATCCAAATGG 609
              |||||||||||||||||||||||||||||||||||||
Db      90611 GTAAGAACTCCCTTATCACTAACCCCCATCCAAATGG 90647


RESULT 7
BC031800
LOCUS       BC031800                673 bp    mRNA    linear   ROD 07-AUG-2002
DEFINITION  Mus musculus, RIKEN cDNA 1810010M01 gene, clone MGC:19113
            IMAGE:4208847, mRNA, complete cds.
ACCESSION   BC031800
VERSION     BC031800.1  GI:21594153
KEYWORDS    MGC.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 673)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUN-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amg@bcm.tmc.edu
            Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
            Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
            Richards, S., Gibbs, R.A.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 24 Row: p Column: 11
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Similarity but not
            identity to protein.
FEATURES             Location/Qualifiers
     source          1. .673
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /map="FVB/N"
                     /clone="MGC:19113 IMAGE:4208847"
                     /tissue_type="Colon, normal. 5 month old male mouse."
                     /clone_lib="NCI_CGAP_Co24"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS             5. .508
                     /codon_start=1
                     /product="RIKEN cDNA 1810010M01 gene"
                     /protein_id="AAH31800.1"
                     /db_xref="GI:21594154"
                     /db_xref="LocusID:69036"
                     /translation="MLAVALLVLLCASASANSIQSRTSSYSGEYGGKGGKRFSHSGNQ
                     LDGPITAFRIRVNRYYIVGLQVRYGTVWSDYVGGTQGDLEEIFLHPGESVIQVSGKYK
                     SYVKQMIFVTDKGRYLPFGKASGTSFNAVPLHPNTVLRFISGRSGSAIDSISLHWDTY
                     PSHCNTC"
BASE COUNT      185 a    165 c    162 g    161 t
ORIGIN

  Query Match             36.8%;  Score 368.8;  DB 10;  Length 673;
  Best Local Similarity   82.9%;  Pred. No. 1.1e-96;
  Matches  421;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy         30 CAGAATGTTGACAGTCGCTCTCCTAGCCCTTCTCTGTGCCTCAGCCTCTGGCAATGCCAT 89
              |||||||||| | ||||| ||| ||| |||||| |||||||| |||||||  ||| ||||
Db          1 CAGAATGTTGGCCGTCGCCCTCTTAGTCCTTCTTTGTGCCTCGGCCTCTGCTAATTCCAT 60

Qy         90 TCAGGCCAGGTCTTCCTCCTATAGTGGAGAGTATGGAAGTGGTGGTGGAAAGCGATTCTC 149
              |||| ||||| | ||||| || ||||||||||||||| ||   || ||||||||||||||
Db         61 TCAGTCCAGGACATCCTCTTACAGTGGAGAGTATGGAGGTAAAGGAGGAAAGCGATTCTC 120

Qy        150 TCATTCTGGCAACCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACACATA 209
              |||||||||||||||| |||||||||||||||||||| ||||  |||| |||||||  |||
Db        121 TCATTCTGGCAACCAGCTGGACGGCCCCATCACCGCCTTCCGTATCCGGGTCAACAGATA 180

Qy        210 CTACATCGTAGGTCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGGTCG 269
              ||||||  |||||||| ||||||||||||||||  ||||||| |||||||  |||||
Db        181 CTACATAGTAGGTCTCCAGGTGCGCTATGGCACAGTGTGGAGTGACTATGTAGGTGGCAC 240

Qy        270 CAACGGAGACCTGGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTCTGG 329
               | |||||||||||||||||||||||||||||| |||||||| |||||||||| |||||
Db        241 ACAGGGAGACCTGGAGGAGATCTTTCTGCACCCCGGGGAATCTGTGATCCAGGTGTCTGG 300

Qy        330 GAAGTACAAGTGGTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCTGTC 389
              ||||||||  |||  ||||| || || | || |||| |||||||||||||| ||| |
Db        301 CAAGTACAAATCATACGTGAAGCAGATGATCTTCGTGACGGACAAGGGCCGCTACCTGCC 360

Qy        390 TTTTGGGAAAGACAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGTGCT 449
              |||||| |||| |   |||||||||||||| || || |||||||| |||||||| || ||
Db        361 TTTTGGAAAAGCCTCAGGCACAAGTTTCAACGCTGTTCCCTTGCATCCCAACACTGTCCT 420

Qy        450 CCGCTTCATCAGTGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGATGT 509
              ||| ||||| ||||| || ||||| |||  ||| ||| | ||| ||||||||||||||
Db        421 CCGTTTCATTAGTGGGCGATCTGGCTCTGCCATAGATTCTATTAGCCTGCACTGGGATAC 480

Qy        510 TTACCCCACTAGCTGCAGCAGATGCTGA 537
              ||||| |    ||||| ||  ||||||
Db        481 CTACCCTAGCCACTGCAACACTTGCTGA 508


RESULT 8
CVU94951/c
LOCUS       CVU94951                3409 bp    DNA     linear   SYN 28-AUG-1998
DEFINITION  Cloning vector pZIP1, complete sequence.
ACCESSION   U94951
VERSION     U94951.1  GI:2218120
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 3409)
  AUTHORS   Kerschbaumer,R.J., Hirschl,S., Kaufmann,A., Ibl,M., Koenig,R. and
            Himmler,G.
  TITLE     Single-chain Fv fusion proteins suitable as coating and detecting
            reagents in a double antibody sandwich enzyme-linked immunosorbent
            assay
  JOURNAL   Anal. Biochem. 249 (2), 219-227 (1997)
  MEDLINE   97356438
   PUBMED   9212874
REFERENCE   2  (bases 1 to 3409)
  AUTHORS   Kerschbaumer,R.J. and Himmler G.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAR-1997) Strainimprovement, Institute of Applied
            Microbiology, Muthgasse 18, Vienna A-1190, Austria
```

```
FEATURES             Location/Qualifiers
     source          1. .3409
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="designed to fuse single-chain Fv to a hinge-region
                     and a leucine-zipper"
     CDS             201. .1061
                     /codon_start=1
                     /transl_table=11
                     /product="beta-lactamase"
                     /protein_id="AAC33185.1"
                     /db_xref="GI:2218121"
                     /translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGY
                     IELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVE
                     YSPVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRL
                     DRWEPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPL
                     LRSALPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIA
                     EIGASLIKHW"
     CDS             2274. .2528
                     /note="pelB-leader is followed by a polylinker"
                     /codon_start=1
                     /transl_table=11
                     /product="hinge-region and leucine-zipper with N-terminal
                     PelB-leader and C-terminal hexahistidine tag"
                     /protein_id="AAC33186.1"
                     /db_xref="GI:2218122"
                     /translation="MKYLLPTAAAGLLLLAAQPAMAQVQLQARLAAAPKPSTPPGSSR
                     MKQLEDKVEELLSKNYHLENEVARLKKLVGERGGHHHHHH"
BASE COUNT      830 a    880 c    840 g    859 t
ORIGIN
```

```
  Query Match             35.7%;  Score 357.2;  DB 12;  Length 3409;
  Best Local Similarity   84.8%;  Pred. No. 3.4e-93;
  Matches  374;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy        561 CTGTGATGAGGAGTAAGAACTCCCTTATCACTAACCCCCATCCAAATGGNNNNNNNNNNN 620
              |||   |  | | || | | | | |||| |||  | ||     | |
Db       2298 CTGCCGTAGGCAATAGGTATTTCATTATGACTGTCTCCTTGAAATAGAATTTTCATGGAA 2239

Qy        621 NNNNNNNNNNNNNNNNNNNNNNNNNNNNTCCTGTGTGAAATTGTTATCCGCTCACAATTC 680
                                          ||||||||||||||||||||||||||||||||
Db       2238 GCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTC 2179

Qy        681 CACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCT 740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2178 CACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCT 2119

Qy        741 AACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCC 800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2118 AACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCC 2059

Qy        801 AGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTT 860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2058 AGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTT 1999

Qy        861 CCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAG 920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1998 CCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAG 1939

Qy        921 CTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATACGCAGGAAAGAACA 980
              |||||||||||||||||||||||||||||||||||||||  |||||||||||||||
Db       1938 CTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACA 1879

Qy        981 TGTGAGCAAAAGGCCAGCAAA 1001
              |||||||||||||||||||||
Db       1878 TGTGAGCAAAAGGCCAGCAAA 1858
```

```
RESULT 9
AF174493/c
LOCUS       AF174493                3595 bp    DNA     circular SYN 29-SEP-2000
DEFINITION  Cloning vector pCL, complete sequence.
ACCESSION   AF174493
VERSION     AF174493.1  GI:5834244
KEYWORDS    .
SOURCE      Cloning vector pCL.
  ORGANISM  Cloning vector pCL
            artificial sequences; vectors.
REFERENCE   1  (bases 1 to 3595)
  AUTHORS   Uhde,K., Kerschbaumer,R.J., Koenig,R., Hirschl,S., Lemaire,O.,
            Boonham,N., Roake,W. and Himmler,G.
  TITLE     Improved detection of Beet necrotic yellow vein virus in a DAS
            ELISA by means of antibody single chain fragments (scFv) which were
            selected to protease-stable epitopes from phage display libraries
  JOURNAL   Arch. Virol. 145 (1), 179-185 (2000)
  MEDLINE   20129515
   PUBMED   10664416
REFERENCE   2  (bases 1 to 3595)
  AUTHORS   Kerschbaumer,R.J., Hirschl,S. and Himmler,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-1999) Strain Improvement, Institute of Applied
            Microbiology, Muthgasse 18, Vienna, AT A-1190, Austria
FEATURES             Location/Qualifiers
     source          1. .3595
                     /organism="Cloning vector pCL"
                     /db_xref="taxon:102687"
                     /note="vector designed to fuse-single chain Fv to human
                     C-kappa domain"
     CDS             201. .1061
                     /note="ampicillin resistance gene"
                     /codon_start=1
                     /transl_table=11
                     /product="beta-lactamase"
                     /protein_id="AAD53875.1"
                     /db_xref="GI:5834245"
                     /translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGY
                     IELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVE
                     YSPVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRL
                     DRWEPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPL
                     LRSALPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIA
                     EIGASLIKHW"
     CDS             2274. .2711
                     /note="contains PelB-leader peptide, polylinker-site,
                     constant domain from human IgG kappa light chain, and
                     hexahistidine tag"
                     /codon_start=1
                     /transl_table=11
                     /product="PelB-IgG kappa light chain fusion protein"
                     /protein_id="AAD53876.1"
                     /db_xref="GI:5834246"
                     /translation="MKYLLPTAAAGLLLLAAQPAMAQVQLQARLAAATVAAPSVFIFP
                     PSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSS
                     TLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGESHHHHHH"
BASE COUNT      881 a    940 c    884 g    890 t
ORIGIN
```

```
  Query Match             35.7%;  Score 357.2;  DB 12;  Length 3595;
  Best Local Similarity   84.8%;  Pred. No. 3.5e-93;
  Matches  374;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy        561 CTGTGATGAGGAGTAAGAACTCCCTTATCACTAACCCCCATCCAAATGGNNNNNNNNNNN 620
              |||   |  | | || | | | | |||| |||  | ||     | |
Db       2298 CTGCCGTAGGCAATAGGTATTTCATTATGACTGTCTCCTTGAAATAGAATTTTCATGGAA 2239

Qy        621 NNNNNNNNNNNNNNNNNNNNNNNNNNNNTCCTGTGTGAAATTGTTATCCGCTCACAATTC 680
                                          ||||||||||||||||||||||||||||||||
Db       2238 GCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTC 2179

Qy        681 CACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCT 740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2178 CACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCT 2119

Qy        741 AACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCC 800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2118 AACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCC 2059
```

```
Qy        801 AGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTT 860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2058 AGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTT 1999

Qy        861 CCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAG 920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1998 CCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAG 1939

Qy        921 CTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGGATACGCAGGAAAGAACA 980
              ||||||||||||||||||||||||||||||||||||||||||||   ||||||||||||||
Db       1938 CTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACA 1879

Qy        981 TGTGAGCAAAAGGCCAGCAAA 1001
              |||||||||||||||||||||
Db       1878 TGTGAGCAAAAGGCCAGCAAA 1858


RESULT 10
AF146395/c
LOCUS       AF146395               3733 bp    DNA     circular SYN 08-JUN-1999
DEFINITION  Cloning vector pCLZip complete sequence.
ACCESSION   AF146395
VERSION     AF146395.1  GI:5006904
KEYWORDS    .
SOURCE      Cloning vector pCLZip.
  ORGANISM  Cloning vector pCLZip
            artificial sequences; vectors.
REFERENCE   1  (bases 1 to 3733)
  AUTHORS   Kerschbaumer,R.J. and Himmler,G.
  TITLE     Dedicated expression vectors for the production of diagnostic
            reagents
  JOURNAL   (in) Harper,K. and Ziegler,A. (Eds.);
            RECOMBINANT ANTIBODIES: APPLICATIONS IN PLANT SCIENCE AND PLANT
            PATHOLOGY: 57-82;
            (1999)
REFERENCE   2  (bases 1 to 3733)
  AUTHORS   Kerschbaumer,R.J., Hirschl,S. and Himmler,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-APR-1999) Strain Improvement, Institute of Applied
            Microbiology, Muthgasse 18, Vienna, AT A-1190, Austria
FEATURES             Location/Qualifiers
     source          1. .3733
                     /organism="Cloning vector pCLZip"
                     /db_xref="taxon:95174"
     gene            201. .1061
                     /gene="bla"
     CDS             201. .1061
                     /gene="bla"
                     /codon_start=1
                     /transl_table=11
                     /product="beta-lactamase"
                     /protein_id="AAD37723.1"
                     /db_xref="GI:5006905"
                     /translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGY
                     IELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVE
                     YSPVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRL
                     DRWEPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPL
                     LRSALPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIA
                     EIGASLIKHW"
     CDS             2274. .2852
                     /note="contains PelB leader region, constant domain from
                     human IgG kappa light chain, hinge-region, leucine-zipper
                     helical region and hexahistidine tag"
                     /codon_start=1
                     /transl_table=11
                     /product="PelB-IgG kappa light chain fusion protein"
                     /protein_id="AAD37724.1"
                     /db_xref="GI:5006906"
                     /translation="MKYLLPTAAAGLLLLAAQPAMAQVQLQARLAAATVAAPSVFIFP
                     PSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSS
                     TLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGESGGPKPSTPPGSSRMKQLEDKV
                     EELLSKNYHLENEVARLKKLVGERGGHHHHHH"
BASE COUNT      917 a    974 c    924 g    918 t
```

```
ORIGIN

  Query Match           35.7%;  Score 357.2;  DB 12;  Length 3733;
  Best Local Similarity  84.8%;  Pred. No. 3.5e-93;
  Matches  374;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy        561 CTGTGATGAGGAGTAAGAACTCCCTTATCACTAACCCCCATCCAAATGGNNNNNNNNNNNN 620
              |||   |  |.| || | | | | |||| |||  | ||     | |
Db       2298 CTGCCGTAGGCAATAGGTATTTCATTATGACTGTCTCCTTGAAATAGAATTTTCATGGAA 2239

Qy        621 NNNNNNNNNNNNNNNNNNNNNNNNNNNNTCCTGTGTGAAATTGTTATCCGCTCACAATTC 680
                                          ||||||||||||||||||||||||||||||||
Db       2238 GCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTC 2179

Qy        681 CACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCT 740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2178 CACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCT 2119

Qy        741 AACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCC 800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2118 AACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCC 2059

Qy        801 AGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTT 860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2058 AGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTT 1999

Qy        861 CCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAG 920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1998 CCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAG 1939

Qy        921 CTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGGATACGCAGGAAAGAACA 980
              ||||||||||||||||||||||||||||||||||||||||||||   ||||||||||||||
Db       1938 CTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACA 1879

Qy        981 TGTGAGCAAAAGGCCAGCAAA 1001
              |||||||||||||||||||||
Db       1878 TGTGAGCAAAAGGCCAGCAAA 1858


RESULT 11
CVU35316/c
LOCUS       CVU35316               4630 bp    DNA     circular SYN 28-AUG-1998
DEFINITION  Cloning vector pDAP2, complete sequence.
ACCESSION   U35316
VERSION     U35316.1  GI:1016008
KEYWORDS    .
SOURCE      Cloning vector pDAP2.
  ORGANISM  Cloning vector pDAP2
            artificial sequences; vectors.
REFERENCE   1  (bases 1 to 4630)
  AUTHORS   Kerschbaumer,R.J., Hirschl,S., Schwager,C., Ibl,M. and Himmler,G.
  TITLE     pDAP2: a vector for construction of alkaline phosphatase
            fusion-proteins
  JOURNAL   Immunotechnology 2 (2), 145-150 (1996)
  MEDLINE   98040653
   PUBMED   9373322
REFERENCE   2  (bases 1 to 4630)
  AUTHORS   Kerschbaumer,R.J., Hirschl,S. and Himmler,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-SEP-1995) Randolf J. Kerschbaumer, Strainimprovement,
            Institute of Applied Microbiology, Nussdorfer Laende 11, Vienna, AT
            A-1190, Austria
FEATURES             Location/Qualifiers
     source          1. .4630
                     /organism="Cloning vector pDAP2"
                     /db_xref="taxon:42878"
                     /plasmid="pDAP2"
                     /note="vector for fusion of proteins or peptides to the
                     N-terminus of E.coli alkaline phosphatase"
     CDS             201. .1061
                     /codon_start=1
                     /transl_table=11
                     /product="beta-lactamase"
```

```
                     /protein_id="AAC33181.1"
                     /db_xref="GI:1016009"
                     /translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGY
                     IELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVE
                     YSPVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRL
                     DRWEPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPL
                     LRSALPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIA
                     EIGASLIKHW"
     CDS             2274. .3749
                     /note="alkaline phosphatase with N-terminal PelB-leader
                     and C-terminal hexa-histidine tag; The coding region for
                     the PelB-leader is followed by a polylinker and the E.coli
                     JM81 alkaline phosphatase gene (ecphoA1-gene)
                     respectively. Additionally the coding sequence for six
                     histidine residues is linked to the 3'end of the
                     ecphoA1-gene"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAC33182.1"
                     /db_xref="GI:3482891"
                     /translation="MKYLLPTAAAGLLLLAAQPAMAQVQLQARLAAAARAPEMPVLEN
                     RAAQGDITAPGGARRLTGDQTAALRDSLSDKPAKNIILLIGDGMGDSEITAARNYAEG
                     AGGFFKGIDALPLTGQYTHYALNKKTGKPDYVTDSAASATAWSTGVKTYNGALGVDIH
                     EKDHPTILEMAKAAGLATGNVSTAELQDATPAALVAHVTSRKCYGPSATSEKCPGNAL
                     EKGGKGSITEQLLNARADVTLGGGAKTFAETATAGEWQGKTLREQAQARGYQLVSDAA
                     SLNSVTEANQQKPLLGLFADGNMPVRWLGPKATYHGNIDKPAVTCTPNPQRNDSVPTL
                     AQMTDKAIELLSKNEKGFFLQVEGASIDKQDHAANPCGQIGETVDLDEAVQRALEFAK
                     KEGNTLVIVTADHAHASQIVAPDTKAPGLTQALNTKDGAVMVMSYGNSEEDSQEHTGS
                     QLRIAAYGPHAANVVGLTDQTDLFYTMKAALGDIAHHHHHH"
BASE COUNT      1146 a    1201 c    1183 g    1100 t
ORIGIN
```

```
  Query Match             35.7%;  Score 357.2;  DB 12;  Length 4630;
  Best Local Similarity   84.8%;  Pred. No. 3.6e-93;
  Matches  374;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy        561 CTGTGATGAGGAGTAAGAACTCCCTTATCACTAACCCCCATCCAAATGGNNNNNNNNNNNN 620
              |||   |  | | || | | | | |||| |||  | ||     | |
Db       2298 CTGCCGTAGGCAATAGGTATTTCATTATGACTGTCTCCTTGAAATAGAATTTTCATGGAA 2239

Qy        621 NNNNNNNNNNNNNNNNNNNNNNNNNNNNTCCTGTGTGAAATTGTTATCCGCTCACAATTC 680
                                          ||||||||||||||||||||||||||||||||
Db       2238 GCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTC 2179

Qy        681 CACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCT 740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2178 CACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCT 2119

Qy        741 AACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCC 800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2118 AACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCC 2059

Qy        801 AGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTT 860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2058 AGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTT 1999

Qy        861 CCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAG 920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1998 CCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAG 1939

Qy        921 CTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGGATACGCAGGAAAGAACA 980
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1938 CTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACA 1879

Qy        981 TGTGAGCAAAAGGCCAGCAAA 1001
              |||||||||||||||||||||
Db       1878 TGTGAGCAAAAGGCCAGCAAA 1858
```

```
RESULT 12
CVU94950/c
LOCUS       CVU94950                4630 bp    DNA     linear   SYN 28-AUG-1998
DEFINITION  Cloning vector pDAP2/S, complete sequence.

ACCESSION   U94950
VERSION     U94950.1  GI:2218117
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 4630)
  AUTHORS   Kerschbaumer,R.J., Hirschl,S., Kaufmann,A., Ibl,M., Koenig,R. and
            Himmler,G.
  TITLE     Single-chain Fv fusion proteins suitable as coating and detecting
            reagents in a double antibody sandwich enzyme-linked immunosorbent
            assay
  JOURNAL   Anal. Biochem. 249 (2), 219-227 (1997)
  MEDLINE   97356438
   PUBMED   9212874
REFERENCE   2  (bases 1 to 4630)
  AUTHORS   Kerschbaumer,R.J. and Himmler G.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAR-1997) Strainimprovement, Institute of Applied
            Microbiology, Muthgasse 18, Vienna A-1190, Austria
FEATURES             Location/Qualifiers
     source          1. .4630
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="designed to fuse proteins to the N-terminus of a
                     modified E.coli alkaline phosphatase with an increased
                     specific activity"
     CDS             201. .1061
                     /codon_start=1
                     /transl_table=11
                     /product="beta-lactamase"
                     /protein_id="AAC33183.1"
                     /db_xref="GI:2218118"
                     /translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGY
                     IELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVE
                     YSPVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRL
                     DRWEPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPL
                     LRSALPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIA
                     EIGASLIKHW"
     CDS             2274. .3749
                     /note="pelB-leader is followed by a polylinker; coding
                     sequence of E.coli JM81 alkaline phosphatase gene mutated
                     to change aspartate in the active site to serine to
                     enhance the specific activity of the enzyme; Mandecki et
                     al., (1991) Protein Eng. 4, 801-804"
                     /codon_start=1
                     /transl_table=11
                     /product="modified E.coli alkaline phosphatase with
                     N-terminal PelB-leader and C-terminal hexahistidine tag"
                     /protein_id="AAC33184.1"
                     /db_xref="GI:2218119"
                     /translation="MKYLLPTAAAGLLLLAAQPAMAQVQLQARLAAAARAPEMPVLEN
                     RAAQGDITAPGGARRLTGDQTAALRDSLSDKPAKNIILLIGDGMGDSEITAARNYAEG
                     AGGFFKGIDALPLTGQYTHYALNKKTGKPDYVTSSAASATAWSTGVKTYNGALGVDIH
                     EKDHPTILEMAKAAGLATGNVSTAELQDATPAALVAHVTSRKCYGPSATSEKCPGNAL
                     EKGGKGSITEQLLNARADVTLGGGAKTFAETATAGEWQGKTLREQAQARGYQLVSDAA
                     SLNSVTEANQQKPLLGLFADGNMPVRWLGPKATYHGNIDKPAVTCTPNPQRNDSVPTL
                     AQMTDKAIELLSKNEKGFFLQVEGASIDKQDHAANPCGQIGETVDLDEAVQRALEFAK
                     KEGNTLVIVTADHAHASQIVAPDTKAPGLTQALNTKDGAVMVMSYGNSEEDSQEHTGS
                     QLRIAAYGPHAANVVGLTDQTDLFYTMKAALGDIAHHHHHH"
BASE COUNT      1145 a    1202 c    1182 g    1101 t
ORIGIN
```

```
  Query Match             35.7%;  Score 357.2;  DB 12;  Length 4630;
  Best Local Similarity   84.8%;  Pred. No. 3.6e-93;
  Matches  374;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy        561 CTGTGATGAGGAGTAAGAACTCCCTTATCACTAACCCCCATCCAAATGGNNNNNNNNNNNN 620
              |||   |  | | || | | | | |||| |||  | ||     | |
Db       2298 CTGCCGTAGGCAATAGGTATTTCATTATGACTGTCTCCTTGAAATAGAATTTTCATGGAA 2239

Qy        621 NNNNNNNNNNNNNNNNNNNNNNNNNNNNTCCTGTGTGAAATTGTTATCCGCTCACAATTC 680
                                          ||||||||||||||||||||||||||||||||
```

```
Db       2238 GCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTC 2179

Qy        681 CACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCT 740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2178 CACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCT 2119

Qy        741 AACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCC 800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2118 AACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCC 2059

Qy        801 AGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTT 860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2058 AGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTT 1999

Qy        861 CCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAG 920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1998 CCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAG 1939

Qy        921 CTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGGATACGCAGGAAAGAACA 980
              ||||||||||||||||||||||||||||||||||||  ||||||||||||||||||||
Db       1938 CTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACA 1879

Qy        981 TGTGAGCAAAAGGCCAGCAAA 1001
              |||||||||||||||||||||
Db       1878 TGTGAGCAAAAGGCCAGCAAA 1858


RESULT 13
AX107161/c
LOCUS       AX107161               5086 bp    DNA     linear   PAT 30-APR-2001
DEFINITION  Sequence 1 from Patent WO0125451.
ACCESSION   AX107161
VERSION     AX107161.1  GI:13922646
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 5086)
  AUTHORS   Hornby,D.P. and Matin,M.M.
  TITLE     Cloning vectors
  JOURNAL   Patent: WO 0125451-A 1 12-APR-2001;
            TRANSGENOMIC, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .5086
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="pMTet1"
BASE COUNT     1304 a   1275 c   1247 g   1260 t
ORIGIN

  Query Match             35.5%;  Score 355.8;  DB 6;  Length 5086;
  Best Local Similarity   80.2%;  Pred. No. 9.3e-93;
  Matches  390;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy        516 CACTAGCTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGATGAGGAGTA 575
              ||||| || ||  |  | | ||  ||    |    |  ||   || | |  |  | | |
Db       3910 CACTAACTTCAATAATTTCAGATGCTTGCTTTTGAGTTAGATCCAATTTACTACGAATCA 3851

Qy        576 AGAACTCCCTTATCACTAACCCCCATCCAAATGGNNNNNNNNNNNNNNNNNNNNNNNNNNN 635
              |   |    ||     | |    |||||    |
Db       3850 ATTTCAATATTTCAGGTTTCATGGATCCATTGCGGCCGCACAGCAGCTTGGCGTAATCAT 3791

Qy        636 NNNNNNNNNNNNNTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAG 695
                          ||||||||||||||||||||||||||||||||||||||||||||||||
Db       3790 GGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAG 3731

Qy        696 CCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTG 755
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3730 CCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTG 3671

Qy        756 CGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAA 815
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3670 CGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAA 3611

Qy        816 TCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCA 875
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3610 TCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCA 3551

Qy        876 CTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGG 935
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3550 CTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGG 3491

Qy        936 TAATACGGTTATCCACAGAATCAGGGGGATACGCAGGAAAGAACATGTGAGCAAAAGGCC 995
              |||||||||||||||||||||||||  ||||||||||||||||||||||||||||||||
Db       3490 TAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCC 3431

Qy        996 AGCAAA 1001
              ||||||
Db       3430 AGCAAA 3425


RESULT 14
AF438204/c
LOCUS       AF438204               5086 bp    DNA     circular SYN 14-NOV-2001
DEFINITION  Positive selection vector pMTet1, complete sequence.
ACCESSION   AF438204
VERSION     AF438204.1  GI:16923909
KEYWORDS    .
SOURCE      Positive selection vector pMTet1.
  ORGANISM  Positive selection vector pMTet1
            artificial sequences; vectors.
REFERENCE   1  (bases 1 to 5086)
  AUTHORS   Matin,M.M. and Hornby,D.P.
  TITLE     A positive selection vector combining tetracycline resistance that
            eliminates the need for bacterial plating
  JOURNAL   Anal. Biochem. 278 (1), 46-51 (2000)
  MEDLINE   20108508
   PUBMED   10640352
REFERENCE   2  (bases 1 to 5086)
  AUTHORS   Matin,M.M. and Hornby,D.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2001) Molecular Biology, Krebs Institute, Firth
            Court, Western Bank, Sheffield S10 2TN, UK
FEATURES             Location/Qualifiers
     source          1. .5086
                     /organism="Positive selection vector pMTet1"
                     /db_xref="taxon:177164"
     gene            92. .1282
                     /gene="tetA(C)"
     CDS             92. .1282
                     /gene="tetA(C)"
                     /note="modified form of the pBR322 tetA(C)"
                     /codon_start=1
                     /transl_table=11
                     /product="tetracycline resistance protein"
                     /protein_id="AAL31631.1"
                     /db_xref="GI:16923910"
                     /translation="MKSNNALIVILGTVTLDAVGIGLVMPVLPGLLRDIVHSDSIASH
                     YGVLLALYALMQFLCAPVLGALSDRFGRRPVLLASLLGATIDYAIMATTPVLWILYAG
                     RIVAGITGATGAVAGAYIADITDGEDRARHFGLMSACFGVGMVAGPVAGGLLGAISLH
                     APFLAAAVLNGLNLLLGCFLMQESHKGERRPMPLRAFNPVSSFRWARGMTIVAALMTV
                     FFIMQLVGQVPAALWVIFGEDRFRWSATMIGLSLAVFGILHALAQAFVTGPATKRFGE
                     KQAIIAGMAADALGYVLLAFATRGWMAFPIMILLASGGIGMPALQAMLSRQVDDDHQG
                     QLQGSLAALTSLTSITGPLIVTAIYAASASTWNGLAWIVGAALYLVCLPALRRGAWSR
                     ATST"
     CDS             3829. .5085
                     /note="from M.MspIm"
                     /codon_start=1
                     /transl_table=11
                     /product="C5 DNA methyltransferase"
                     /protein_id="AAL31632.1"
                     /db_xref="GI:16923911"
                     /translation="MKPEILKLIRSKLDLTQKQASEIIEVSDKAWQQWESGKTEMHPA
                     YYSFLQEKLKDKINFEELSAQKTLQKKIFDKYNQNQITKNAEELAEITHIEERKDAYS
                     SDFKFIDLFSGIGGIRQSFEVNGGKCVFSSEIDPFAKFTYYTNFGVVPFGDITKVEAT
                     TIPEHDILCAGFPGQPWSHMGKREGFEHPTQGTMFHEIVRIIETKKTPVLFLENVPGL
```

```
                    INHDDGNTLKVIIETLEDMGYKVHHTVLDASHFGIPQKRKRFYLVAFLNQNIHFEFPK
                    PPMISKDIGEVLESDVTGYSISEHLQKSYLFKKDDGKPSLIDKNTTGAVKTLVSTYHK
                    IQRLTGTFVKDGETAIRLLTTNECKAIMGFPKDFVIPVSRTQMYRQMGNSVVVPVVTK
                    IAEQISLALKTVNQQSPQENFELELV"
BASE COUNT      1304 a    1275 c    1248 g    1259 t
ORIGIN

  Query Match             35.5%;  Score 355.8;  DB 12;  Length 5086;
  Best Local Similarity   80.2%;  Pred. No. 9.3e-93;
  Matches  390;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy        516 CACTAGCTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGATGAGGAGTA 575
              ||||| || ||  |  | | ||  ||    |   | ||   || | |  |  | |  |
Db       3910 CACTAACTTCAATAATTTCAGATGCTTGCTTTTGAGTTAGATCCAATTTACTACGAATCA 3851

Qy        576 AGAACTCCCTTATCACTAACCCCCATCCAAATGGNNNNNNNNNNNNNNNNNNNNNNNNNN 635
              |   |    ||     | |    |||||   |
Db       3850 ATTTCAATATTTCAGGTTTCATGGATCCATTGCGGCCGCACAGCAGCTTGGCGTAATCAT 3791

Qy        636 NNNNNNNNNNNNNTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAG 3791
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3790 GGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAG 695

Qy        696 CCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTG 3731
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3730 CCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTG 755

Qy        756 CGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAA 3671
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3670 CGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAA 815

Qy        816 TCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCA 3611
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3610 TCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCA 875

Qy        876 CTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGG 3551
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3550 CTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGG 935

Qy        936 TAATACGGTTATCCACAGAATCAGGGGGATACGCAGGAAAGAACATGTGAGCAAAAGGCC 3491
              ||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||||
Db       3490 TAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCC 995

Qy        996 AGCAAA 1001
              ||||||
Db       3430 AGCAAA 3425


RESULT 15
AF153422
LOCUS       AF153422                3417 bp    DNA     circular SYN 02-APR-2000
DEFINITION  Cloning vector pTG8, complete sequence.
ACCESSION   AF153422
VERSION     AF153422.1  GI:7384995
KEYWORDS    .
SOURCE      Cloning vector pTG8.
  ORGANISM  Cloning vector pTG8
            artificial sequences; vectors.
REFERENCE   1  (bases 1 to 3417)
  AUTHORS   Yang,Y. and Spector,A.
  TITLE     Improved cloning vectors for transgene construction
  JOURNAL   BioTechniques 22 (6), 1032-1034 (1997)
  MEDLINE   97331414
  PUBMED    9187746
REFERENCE   2  (bases 1 to 3417)
  AUTHORS   Yang,Y. and Spector,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAY-1999) Institute of Molecular Biology, University
            of Hong Kong, 8 Sassoon, Pokfulam, Hong Kong
FEATURES             Location/Qualifiers
     source          1. .3417
                     /organism="Cloning vector pTG8"
                     /db_xref="taxon:119062"
                     /note="derived from Strategene vector pBS KS(+) by
                     insertion of rare FseI and PacI restriction sites flanking
     CDS             the multiple cloning site"
                     complement(241. .834)
                     /codon_start=1
                     /transl_table=11
                     /product="lacZ"
                     /protein_id="AAF61634.1"
                     /db_xref="GI:7384996"
                     /translation="MTMITPSAQAGQLTLTKGNKSWVPGPPSRSTVSISLISNSCSPG
                     DPLVLERPPPRWSSNSPYSESYYLINARSLAVVLQRRDWENPGVTQLNRLAAHPPFAS
                     WRNSEEARTDRPSQQLRSLNGEWDAPCSGALSAAGVVVTRSVTATLASALAPAPFAFF
     CDS             PSFLATFAGFPRQALNRGLPLGFRFSALRHLDPKKLD"
                     complement(2316. .2867)
                     /codon_start=1
                     /transl_table=11
                     /product="chloramphenicol acetyltransferase"
                     /protein_id="AAF61635.1"
                     /db_xref="GI:7384997"
                     /translation="MEKKITGYTTVDISQWHRKEHFEAFQSVAQCTYNQTVQLDITAF
                     LKTVKKNKHKFYPAFIHILARLMNAHPELRMAMKDGELVIWDSVHPCYTVFHEQTETF
                     SSLWSEYHDDFRQFLHIYSQDVACYGENLAYFPKGFIENMFFVSANPWVSFTSFDLNV
                     ANMDNFFAPFSPWANIIRKATRC"
BASE COUNT      905 a     837 c     837 g     838 t
ORIGIN

  Query Match             35.2%;  Score 352.6;  DB 12;  Length 3417;
  Best Local Similarity   89.2%;  Pred. No. 7.6e-92;
  Matches  355;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy        604 AAATGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCCTGTGTGAAATTG 663
              || |||                                              |||||||||||||||
Db        797 AATTGGCCGGCCTGCGCGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTG 663

Qy        664 TTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGG 723
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        857 TTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGG 723

Qy        724 TGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTC 916
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        917 TGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTC 783

Qy        784 GGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTT 843
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        977 GGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTT 1036

Qy        844 GCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCT 903
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1037 GCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCT 1096

Qy        904 GCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGG 963
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1097 GCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGG 963

Qy        964 ATACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1001
              |||||||||||||||||||||||||||||||||||||||
Db       1157 TAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1194


Search completed: July 16, 2003, 13:27:48
Job time : 1825.86 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 16, 2003, 03:41:41 ; Search time 170.989 Seconds
(without alignments)
13183.634 Million cell updates/sec

Title: US-09-981-353-24
Perfect score: 1001
Sequence: 1 atgattacgaattcgagctc.....gtgagcaaaaggccagcaaa 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	583	58.2	645	19	AAV63528	Consensus sequence
2	583	58.2	665	20	AAX28677	Clone HP01738 enco
3	583	58.2	739	21	AAZ45659	DNA sequence of co
4	579.8	57.9	718	19	AAV63527	Full length sequen
5	545.8	54.5	560	18	AAT45885	Human colon specif
6	545.8	54.5	560	19	AAV16673	Polynucleotide seq
7	545.8	54.5	560	24	ABK15461	Human colon specif
8	501	50.0	501	20	AAX28676	Clone HP01738 enco
9	429.4	42.9	661	22	AAK88295	Human digestive sy
c 10	355.8	35.5	5086	22	AAF84674	Nucleotide sequenc
c 11	352.4	35.2	5924	21	AAZ38921	hCAT1 binding huma
c 12	352.4	35.2	5925	21	AAZ38770	HCAT1 clone 25 ant
13	352.2	35.2	3789	21	AAZ39782	Plasmid pGB1MMN nu
14	352.2	35.2	5393	14	AAQ34611	Plasmid pR3. Synt
15	352	35.2	3983	22	AAD10265	pUC/Ac vector DNA.
16	351.8	35.1	4824	17	AAT08975	Shortened C1 gene
17	351.8	35.1	4824	20	AAX15634	Nucleotide sequenc
c 18	351.8	35.1	6843	20	AAV83538	PCR-generated regi
c 19	351.8	35.1	7107	20	AAV83539	PCR-generated regi
c 20	351.8	35.1	8710	16	AAQ89650	pSC11 FIPV E1. Sy
c 21	351.8	35.1	9020	16	AAT07652	pSC11 FIPV N. Syn
c 22	351.8	35.1	12011	20	AAV83540	PCR-generated regi
23	351.4	35.1	8062	22	AAD04946	Plasmid pRK76 used
24	351.4	35.1	8153	22	AAD04945	Plasmid pRK74 used
25	351.2	35.1	7958	24	AAI72490	H6-promoter human
c 26	351	35.1	8800	24	ABK95584	Yeast reporter con
c 27	351	35.1	10042	24	ABK95585	Yeast reporter con
28	350.8	35.0	584	22	AAF68704	Human lung tumour
29	350.8	35.0	584	24	ABK38615	cDNA encoding clon
30	350.8	35.0	655	24	ABK27689	Human colon cancer
31	350.8	35.0	738	20	AAZ09716	Plasmid pGEM DNA f
32	350.8	35.0	742	22	AAH82483	Human ovarian tumo
33	350.8	35.0	972	21	AAC83524	T7 promoter sequen
c 34	350.8	35.0	2571	21	AAC62824	Glyceraldehyde-3-p
c 35	350.8	35.0	2571	24	ABK17089	Eucalyptus grandis
36	350.8	35.0	3018	20	AAX29905	Plasmid pGEM (RTM)
37	350.8	35.0	3216	21	AAZ49993	Plasmid pGN1 for e
38	350.8	35.0	3216	22	AAS06382	Vector pGN1 DNA se
39	350.8	35.0	3253	19	AAV43439	DNA seqeunce of th
40	350.8	35.0	3323	24	AAD27061	Plasmid pGN1 DNA.
41	350.8	35.0	3675	24	ABK12468	Plasmid vector pGX
42	350.8	35.0	3715	24	AAD27064	Plasmid TopoRNAi D
43	350.8	35.0	3774	24	AAD27062	Plasmid pGN29 DNA.
44	350.8	35.0	3968	22	AAD09981	pHSP70-1MCS constr
45	350.8	35.0	4001	24	AAD27066	Plasmid pGN59A DNA

ALIGNMENTS

RESULT 1
AAV63528
ID AAV63528 standard; cDNA; 645 BP.
XX
AC AAV63528;
XX
DT 28-JAN-1999 (first entry)
XX
DE Consensus sequence of CS110 sequence.
XX
KW EST sequence; CS110; gastrointestinal tract; cancer; adenocarcinoma;
KW lymphoma; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 44..547
FT /*tag= a
XX
PN WO9844160-A1.
XX
PD 08-OCT-1998.
XX
PF 31-MAR-1998; 98WO-US06338.
XX
PR 31-MAR-1997; 97US-0829755.
XX
PA (ABBO) ABBOTT LAB.
XX
PI Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN;
PI Gordon J, Granados EN, Hayden M, Hodges SC, Klass MR;
PI Kratochvil JD, Roberts-Rapp L, Russell JC, Stroupe SD;

```
XX
DR   WPI; 1998-557139/47.
DR   P-PSDB; AAW80477.
XX
PT   New CS110 polypeptide(s)  - useful for detecting gastrointestinal
PT   tract diseases, e.g. gastrointestinal tract cancers and to produce
PT   antibodies
XX
PS   Claim 1; Page 91; 116pp; English.
XX
CC   The present sequence is a consensus sequence of ESTs AAV63521-27
CC   which represent a set of contiguous and partially overlapping
CC   sequences designated CS110. The sequences are isolated from a CDNA
CC   library made from gastrointestinal tract tumour and normal tissues.
CC   The polypeptides are useful diagnostically to detect CS110
CC   antigen/anti-CS110 antibody in samples, e.g. to detect diseases
CC   and conditions of the gastrointestinal tract, especially
CC   cancers, e.g. adenocarcinoma and lymphoma.
XX
SQ   Sequence 645 BP; 139 A; 174 C; 172 G; 160 T; 0 other;

  Query Match             58.2%;  Score 583;  DB 19;  Length 645;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-130;
  Matches  583;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 CCCCAGAATGTTGACAGTCGCTCTCCTAGCCCTTCTCTGTGCCTCAGCCTCTGGCAATGC 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         37 CCCCAGAATGTTGACAGTCGCTCTCCTAGCCCTTCTCTGTGCCTCAGCCTCTGGCAATGC 96

Qy         87 CATTCAGGCCAGGTCTTCCTCCTATAGTGGAGAGTATGGAAGTGGTGGTGGAAAGCGATT 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         97 CATTCAGGCCAGGTCTTCCTCCTATAGTGGAGAGTATGGAAGTGGTGGTGGAAAGCGATT 156

Qy        147 CTCTCATTCTGGCAACCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACAC 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        157 CTCTCATTCTGGCAACCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACAC 216

Qy        207 ATACTACATCGTAGGTCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGG 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        217 ATACTACATCGTAGGTCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGG 276

Qy        267 TCGCAACGGAGACCTGGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTC 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        277 TCGCAACGGAGACCTGGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTC 336

Qy        327 TGGGAAGTACAAGTGGTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCT 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        337 TGGGAAGTACAAGTGGTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCT 396

Qy        387 GTCTTTTGGGAAAGACAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGT 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        397 GTCTTTTGGGAAAGACAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGT 456

Qy        447 GCTCCGCTTCATCAGTGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGA 506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        457 GCTCCGCTTCATCAGTGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGA 516

Qy        507 TGTTTACCCCACTAGCTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGA 566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        517 TGTTTACCCCACTAGCTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGA 576

Qy        567 TGAGGAGTAAGAACTCCCTTATCACTAACCCCCATCCAAATGG 609
              |||||||||||||||||||||||||||||||||||||||||||
Db        577 TGAGGAGTAAGAACTCCCTTATCACTAACCCCCATCCAAATGG 619


RESULT 2
AAX28677
ID   AAX28677 standard; cDNA; 665 BP.
XX
AC   AAX28677;
XX
DT   26-AUG-1999  (first entry)
XX
DE   Clone HP01738 encoding a human secretory signal protein (2).
XX
KW   Human; secretory signal protein sequence; cell membrane;
KW   proliferation; differentiation; carcinostatic agent; antigen;
KW   antibody; probe; hybridisation; gene therapy; HP01738; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             58..562
FT                   /*tag= a
FT                   /product= "secretory signal protein"
XX
PN   WO9918204-A2.
XX
PD   15-APR-1999.
XX
PF   05-OCT-1998;   98WO-JP04476.
XX
PR   08-OCT-1997;   97JP-0276268.
XX
PA   (PROT-) PROTEGENE INC.
PA   (SAGA ) SAGAMI CHEM RES CENT.
XX
PI   Kato S,  Kobayashi M,  Sekine S,  Yamaguchi T;
XX
DR   WPI; 1999-264020/22.
DR   P-PSDB; AAY03235.
XX
PT   Human proteins with secretory signal sequences and nucleotide
PT   sequences
XX
PS   Claim 4; Page 73-74; 84pp; English.
XX
CC   This is the nucleotide sequence of a clone encoding a human
CC   secretory signal protein sequence, used in the method of the
CC   invention. All of the proteins exist in the cell membrane, so are
CC   considered to be proteins controlling the proliferation and
CC   differentiation of the cells. They may be useful as carcinostatic
CC   agents or as antigens for preparing antibodies against the proteins.
CC   The cDNAs can be used as probes for gene diagnosis and gene sources
CC   for gene therapy, as well as for large-scale expression of the
CC   proteins.
XX
SQ   Sequence 665 BP; 151 A; 176 C; 175 G; 163 T; 0 other;

  Query Match             58.2%;  Score 583;  DB 20;  Length 665;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-130;
  Matches  583;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 CCCCAGAATGTTGACAGTCGCTCTCCTAGCCCTTCTCTGTGCCTCAGCCTCTGGCAATGC 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         52 CCCCAGAATGTTGACAGTCGCTCTCCTAGCCCTTCTCTGTGCCTCAGCCTCTGGCAATGC 111

Qy         87 CATTCAGGCCAGGTCTTCCTCCTATAGTGGAGAGTATGGAAGTGGTGGTGGAAAGCGATT 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        112 CATTCAGGCCAGGTCTTCCTCCTATAGTGGAGAGTATGGAAGTGGTGGTGGAAAGCGATT 171

Qy        147 CTCTCATTCTGGCAACCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACAC 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        172 CTCTCATTCTGGCAACCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACAC 231

Qy        207 ATACTACATCGTAGGTCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGG 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        232 ATACTACATCGTAGGTCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGG 291

Qy        267 TCGCAACGGAGACCTGGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTC 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        292 TCGCAACGGAGACCTGGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTC 351

Qy        327 TGGGAAGTACAAGTGGTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCT 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        352 TGGGAAGTACAAGTGGTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCT 411

Qy        387 GTCTTTTGGGAAAGACAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGT 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        412 GTCTTTTGGGAAAGACAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGT 471

Qy        447 GCTCCGCTTCATCAGTGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGA 506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        472 GCTCCGCTTCATCAGTGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGA 531

Qy        507 TGTTTACCCCACTAGCTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGA 566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        532 TGTTTACCCCACTAGCTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGA 591

Qy        567 TGAGGAGTAAGAACTCCCTTATCACTAACCCCCATCCAAATGG 609
              |||||||||||||||||||||||||||||||||||||||||||
Db        592 TGAGGAGTAAGAACTCCCTTATCACTAACCCCCATCCAAATGG 634
```

RESULT 3
AAZ45659
ID AAZ45659 standard; DNA; 739 BP.
XX
AC AAZ45659;
XX
DT 06-APR-2000 (first entry)
XX
DE DNA sequence of colon specific gene (CSG) clone 1285234.
XX
KW Human; colon specific gene; CSG; colon; metastasis; colon cancer;
KW stomach cancer; small intestine cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9960161-A1.
XX
PD 25-NOV-1999.
XX
PF 12-MAY-1999; 99WO-US10498.
XX
PR 21-MAY-1998; 98US-0086266.
XX
PA (DIAD-) DIADEXUS LLC.
XX
PI Macina RA, Yang F, Sun Y;
XX
DR WPI; 2000-126383/11.
XX
PT Diagnosing, monitoring and staging colon cancer -
XX
PS Example 1; Page 37; 47pp; English.
XX
CC The present sequence represents a clone of a human colon specific
CC gene (CSG) (gene ID number 239588). mRNA expression of the CSGs is
CC highly specific for the colon. The specification describes a method
CC for diagnosing the presence, or metastasis, of colon cancer in a
CC patient. The method comprises measuring CSG levels in a cell, tissue
CC or bodily fluid sample of the patient and a control (i.e. a human
CC with no cancer), where increased CSG levels in the patient compared
CC to the control is associated with the presence, or metastasis, of
CC colon cancer. The methods may also be used for staging colon cancer,
CC or for monitoring colon cancers, particularly of the colon, stomach
CC and small intestine.
XX
SQ Sequence 739 BP; 167 A; 190 C; 183 G; 177 T; 22 other;

Query Match 58.2%; Score 583; DB 21; Length 739;
Best Local Similarity 100.0%; Pred. No. 5.4e-130;
Matches 583; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         27 CCCCAGAATGTTGACAGTCGCTCTCCTAGCCCTTCTCTGTGCCTCAGCCTCTGGCAATGC 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         80 CCCCAGAATGTTGACAGTCGCTCTCCTAGCCCTTCTCTGTGCCTCAGCCTCTGGCAATGC 139

Qy         87 CATTCAGGCCAGGTCTTCCTCCTATAGTGGAGAGTATGGAAGTGGTGGTGGAAAGCGATT 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        140 CATTCAGGCCAGGTCTTCCTCCTATAGTGGAGAGTATGGAAGTGGTGGTGGAAAGCGATT 199

Qy        147 CTCTCATTCTGGCAACCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACAC 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        200 CTCTCATTCTGGCAACCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACAC 259

Qy        207 ATACTACATCGTAGGTCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGG 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        260 ATACTACATCGTAGGTCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGG 319

Qy        267 TCGCAACGGAGACCTGGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTC 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        320 TCGCAACGGAGACCTGGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTC 379

Qy        327 TGGGAAGTACAAGTGGTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCT 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        380 TGGGAAGTACAAGTGGTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCT 439

Qy        387 GTCTTTTGGGAAAGACAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGT 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        440 GTCTTTTGGGAAAGACAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGT 499

Qy        447 GCTCCGCTTCATCAGTGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGA 506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        500 GCTCCGCTTCATCAGTGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGA 559

Qy        507 TGTTTACCCCACTAGCTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGA 566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        560 TGTTTACCCCACTAGCTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGA 619

Qy        567 TGAGGAGTAAGAACTCCCTTATCACTAACCCCCATCCAAATGG 609
              |||||||||||||||||||||||||||||||||||||||||||
Db        620 TGAGGAGTAAGAACTCCCTTATCACTAACCCCCATCCAAATGG 662
```

RESULT 4
AAV63527
ID AAV63527 standard; cDNA; 718 BP.
XX
AC AAV63527;
XX
DT 28-JAN-1999 (first entry)
XX
DE Full length sequence of clone 609120IH corresponding to CS110 sequence.
XX
KW EST sequence; CS110; gastrointestinal tract; cancer; adenocarcinoma;
KW lymphoma; ss.
XX
OS Homo sapiens.
XX
PN WO9844160-A1.
XX
PD 08-OCT-1998.
XX
PF 31-MAR-1998; 98WO-US06338.
XX
PR 31-MAR-1997; 97US-0829755.
XX
PA (ABBO) ABBOTT LAB.
XX
PI Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN;
PI Gordon J, Granados EN, Hayden M, Hodges SC, Klass MR;
PI Kratochvil JD, Roberts-Rapp L, Russell JC, Stroupe SD;
XX
DR WPI; 1998-557139/47.
XX
PT New CS110 polypeptide(s) - useful for detecting gastrointestinal
PT tract diseases, e.g. gastrointestinal tract cancers and to produce
PT antibodies
XX

```
PS   Claim 1; Page 91; 116pp; English.
XX
CC   AAV63521-27 represent a set of contiguous and partially overlapping EST
CC   sequences designated CS110. The sequences are isolated from a cDNA
CC   library made from gastrointestinal tract tumour and normal tissues.
CC   The polypeptides are useful diagnostically to detect CS110
CC   antigen/anti-CS110 antibody in samples, e.g. to detect diseases
CC   and conditions of the gastrointestinal tract, especially
CC   cancers, e.g. adenocarcinoma and lymphoma.
XX
SQ   Sequence 718 BP; 154 A; 202 C; 186 G; 176 T; 0 other;

  Query Match            57.9%;  Score 579.8;  DB 19;  Length 718;
  Best Local Similarity  99.7%;  Pred. No. 3.1e-129;
  Matches  581;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         27 CCCCAGAATGTTGACAGTCGCTCTCCTAGCCCTTCTCTGTGCCTCAGCCTCTGGCAATGC 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        109 CCCCAGAATGTTGACAGTCGCTCTCCTAGCCCTTCTCTGTGCCTCAGCCTCTGGCAATGC 168

Qy         87 CATTCAGGCCAGGTCTTCCTCCTATAGTGGAGAGTATGGAAGTGGTGGTGGAAAGCGATT 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        169 CATTCAGGCCAGGTCTTCCTCCTATAGTGGAGAGTATGGAAGTGGTGGTGGAAAGCGATT 228

Qy        147 CTCTCATTCTGGCAACCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACAC 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        229 CTCTCATTCTGGCAACCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACAC 288

Qy        207 ATACTACATCGTAGGTCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGG 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        289 ATACTACATCGTAGGTCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGG 348

Qy        267 TCGCAACGGAGACCTGGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTC 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        349 TCGCAACGGAGACCTGGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTC 408

Qy        327 TGGGAAGTACAAGTGGTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCT 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        409 TGGGAAGTACAAGTGGTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCT 468

Qy        387 GTCTTTTGGGAAAGACAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGT 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        469 GTCTTTTGGGAAAGACAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGT 528

Qy        447 GCTCCGCTTCATCAGTGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGA 506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        529 GCTCCGCTTCATCAGTGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGA 588

Qy        507 TGTTTACCCCACTAGCTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGA 566
              ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        589 TGTTTACCCCACTAGCTGCAGCAGATGCTGAGCCTCCTCTCCTAGGCAGGGGCACTGTGA 648

Qy        567 TGAGGAGTAAGAACTCCCTTATCACTAACCCCCATCCAAATGG 609
              ||||||||||||||||||||||||||||||||||||| ||||
Db        649 TGAGGAGTAAGAACTCCCTTATCACTAACCCCCATCCATATGG 691


RESULT 5
AAT45885
ID   AAT45885 standard; cDNA; 560 BP.
XX
AC   AAT45885;
XX
DT   13-MAR-1997  (first entry)
XX
DE   Human colon specific gene CSG6 cDNA partial clone.
XX
KW   Colon specific gene; CSG6; colon cancer; metastasis; diagnosis;
KW   gene therapy; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             2..496
FT                   /*tag=  a
XX
PN   WO9639419-A1.
XX
PD   12-DEC-1996.
XX
PF   06-JUN-1995;   95WO-US07289.
XX
PR   06-JUN-1995;   95WO-US07289.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Yu G;
XX
DR   WPI; 1997-043054/04.
DR   P-PSDB; AAW06549.
XX
PT   Human colon specific genes and their expression products - detection
PT   of which, in non-colon tissue samples, can be used as indication of
PT   colon cancer metastasis
XX
PS   Claim 1; Fig 6; 60pp; English.
XX
CC   13 cDNA clones (AAT45880-92), 11 of them partial clones, correspond to
CC   human colon specific genes, designated CSG1, CSG2, etc., that are
CC   primarily expressed in tissues derived from the colon.  CSG7 and
CC   CSG10 show reduced expression in colon cancer cells as compared to
CC   that in normal cells; the remaining genes are overexpressed in
CC   colon cancer.  The partial cDNA sequences can be used to isolate
CC   full-length clones and genomic clones including the complete gene.
CC   CSG nucleic acids can be used to produce CSG polypeptides (see also
CC   AAW06545-53) in transformed host cells, as probes to detect disorders
CC   of the colon, partic. colon cancer and colon cancer metastasis, and
CC   in gene therapy.
XX
SQ   Sequence 560 BP; 113 A; 157 C; 152 G; 138 T; 0 other;

  Query Match            54.5%;  Score 545.8;  DB 18;  Length 560;
  Best Local Similarity  99.5%;  Pred. No. 4.1e-121;
  Matches  558;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy         42 AGTCGCTCTCCTAGCCCTTCTCTGTGCCTCAGCCTCTGGCAATGCCATTCAGGCCAGGTC 101
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db          1 AGTCGCTCTCCTAGCCCTTCTCTGTGCCTCACCCTCTGGCAATGCCATTCAGGCCAGGTC 60

Qy        102 TTCCTCCTATAGTGGAGAGTATGGAAGTGGTGGTGGAAAGCGATTCTCTCATTCTGGCAA 161
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db         61 TTCCTCCTATAGTGGAGAGTATGGAGGTGGTGGTGGAAAGCGATTCTCTCATTCTGGCAA 120

Qy        162 CCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACACATACTACATCGTAGG 221
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACACATACTACATCGTAGG 180

Qy        222 TCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGGTCGCAACGGAGACCT 281
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGGTCGCAACGGAGACCT 240

Qy        282 GGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTCTGGGAAGTACAAGTG 341
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTCTGGGAAGTACAAGTG 300

Qy        342 GTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCTGTCTTTTGGGAAAGA 401
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCTGTCTTTTGGGAAAGA 360

Qy        402 CAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGTGCTCCGCTTCATCAG 461
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGTGCTCCGCTTCATCAG 420

Qy        462 TGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGATGTTTACCCCACTAG 521
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGATGTTTACCCCACTAG 480
```

```
Qy        522 CTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGATGAGGAGTAAGAACT 581
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGATGAGGAGTAAGAACT 540

Qy        582 CCCTTATCACTAACCCCCATC 602
                |||||||||||||||||||
Db        541 -CCTTATCACTAACCCCCATC 560


RESULT 6
AAV16673
ID   AAV16673 standard; cDNA; 560 BP.
XX
AC   AAV16673;
XX
DT   22-JUN-1998  (first entry)
XX
DE   Polynucleotide sequence of a colon-specific gene.
XX
KW   Colon-specific gene; probe; detection; expression; human;
KW   diagnostic assay; colon cancer; antibody; screening; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             2..490
FT                   /*tag= a
FT                   /note= "no stop codon given"
XX
PN   US5733748-A.
XX
PD   31-MAR-1998.
XX
PF   06-JUN-1995;   95US-0469667.
XX
PR   06-JUN-1995;   95US-0469667.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen C,  Yu G;
XX
DR   WPI; 1998-229823/20.
DR   P-PSDB; AAW46880.
XX
PT   Colon-specific nucleic acids - useful as probes for detecting colon
PT   cancer micrometastases
XX
PS   Claim 15; Fig 6; 51pp; English.
XX
CC   AAV16668-81 represent polynucleotide sequences of partial or full-length
CC   cDNA clones of colon-specific genes. The polynucleotides can be used
CC   as probes to detect expression of the corresponding human genes, e.g. in
CC   diagnostic assays for detecting micrometastases of colon cancer.
CC   Recombinant cells containing the polynucleotides can be used to
CC   produce the polypeptides, in order that antibodies can be raised and
CC   used in further screening or diagnostics.
XX
SQ   Sequence 560 BP; 113 A; 157 C; 152 G; 138 T; 0 other;

  Query Match           54.5%;  Score 545.8;  DB 19;  Length 560;
  Best Local Similarity  99.5%;  Pred. No. 4.1e-121;
  Matches  558;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy         42 AGTCGCTCTCCTAGCCCTTCTCTGTGCCTCAGCCTCTGGCAATGCCATTCAGGCCAGGTC 101
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db          1 AGTCGCTCTCCTAGCCCTTCTCTGTGCCTCACCCTCTGGCAATGCCATTCAGGCCAGGTC 60

Qy        102 TTCCTCCTATAGTGGAGAGTATGGAAGTGGTGGTGGAAAGCGATTCTCTCATTCTGGCAA 161
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db         61 TTCCTCCTATAGTGGAGAGTATGGAGGTGGTGGTGGAAAGCGATTCTCTCATTCTGGCAA 120

Qy        162 CCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACACATACTACATCGTAGG 221
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACACATACTACATCGTAGG 180

Qy        222 TCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGGTCGCAACGGAGACCT 281
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGGTCGCAACGGAGACCT 240

Qy        282 GGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTCTGGGAAGTACAAGTG 341
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTCTGGGAAGTACAAGTG 300

Qy        342 GTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCTGTCTTTTGGGAAAGA 401
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCTGTCTTTTGGGAAAGA 360

Qy        402 CAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGTGCTCCGCTTCATCAG 461
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGTGCTCCGCTTCATCAG 420

Qy        462 TGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGATGTTTACCCCACTAG 521
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGATGTTTACCCCACTAG 480

Qy        522 CTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGATGAGGAGTAAGAACT 581
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGATGAGGAGTAAGAACT 540

Qy        582 CCCTTATCACTAACCCCCATC 602
                |||||||||||||||||||
Db        541 -CCTTATCACTAACCCCCATC 560


RESULT 7
ABK15461
ID   ABK15461 standard; cDNA; 560 BP.
XX
AC   ABK15461;
XX
DT   08-MAY-2002  (first entry)
XX
DE   Human colon specific gene 6 (CSG6) cDNA.
XX
KW   Human; colon specific gene; CSG; cytostatic; colon disorder; gene;
KW   gastric cancer; ulcerative colitis; granulomatous colitis; colon cancer;
KW   cancer cell; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             2..490
FT                   /*tag= a
FT                   /product= "CSG6 protein"
FT                   /partial
FT                   /note= "No stop codon shown"
XX
PN   US6337195-B1.
XX
PD   08-JAN-2002.
XX
PF   31-MAR-1998;   98US-0224110.
XX
PR   06-JUN-1995;   95US-0469667.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Yu G,  Rosen C;
XX
DR   WPI; 2002-163239/21.
DR   P-PSDB; AAU76209.
XX
PT   Human colon specific gene polypeptide, useful as diagnostic marker,
PT   vaccine and for screening agonists and antagonists for treating colon
PT   cancer -
```

```
XX
PS   Example 1; Fig 6; 49pp; English.
XX
CC   This invention relates to the cDNA and protein sequences of novel
CC   human colon specific genes. These genes which were isolated from a human
CC   colon tissue cDNA library may have cytostatic activity and may be used
CC   as a vaccine against colon disorders such as gastric cancer, ulcerative
CC   colitis and granulomatous colitis. The nucleotide sequences of the
CC   invention may be used to identify corresponding full length gene
CC   sequences and is useful as diagnostic marker for colon cancer and as a
CC   colon cancer vaccine. The protein sequences of the invention are
CC   useful for preparing monoclonal antibodies specific to CSG proteins,
CC   these antibodies can be used to isolate the protein from colon tissue or
CC   for protein quantitation from a sample using enzyme linked immunosorbant
CC   assay (ELISA). It is also useful for targeting cancer cells and for
CC   screening agonist and antagonist for the CSG proteins which are useful
CC   for treating colon cancer. The present sequence represents the human
CC   colon specific gene sequence (CSG6) cDNA sequence of the invention.
XX
SQ   Sequence 560 BP; 113 A; 157 C; 152 G; 138 T; 0 other;

  Query Match             54.5%;  Score 545.8;  DB 24;  Length 560;
  Best Local Similarity   99.5%;  Pred. No. 4.1e-121;
  Matches  558;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy         42 AGTCGCTCTCCTAGCCCTTCTCTGTGCCTCAGCCTCTGGCAATGCCATTCAGGCCAGGTC 101
               ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db          1 AGTCGCTCTCCTAGCCCTTCTCTGTGCCTCACCCTCTGGCAATGCCATTCAGGCCAGGTC 60

Qy        102 TTCCTCCTATAGTGGAGAGTATGGAAGTGGTGGTGGAAAGCGATTCTCTCATTCTGGCAA 161
               |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db         61 TTCCTCCTATAGTGGAGAGTATGGAGGTGGTGGTGGAAAGCGATTCTCTCATTCTGGCAA 120

Qy        162 CCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACACATACTACATCGTAGG 221
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACACATACTACATCGTAGG 180

Qy        222 TCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGGTCGCAACGGAGACCT 281
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGGTCGCAACGGAGACCT 240

Qy        282 GGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTCTGGGAAGTACAAGTG 341
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTCTGGGAAGTACAAGTG 300

Qy        342 GTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCTGTCTTTTGGGAAAGA 401
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCTGTCTTTTGGGAAAGA 360

Qy        402 CAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGTGCTCCGCTTCATCAG 461
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGTGCTCCGCTTCATCAG 420

Qy        462 TGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGATGTTTACCCCACTAG 521
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGATGTTTACCCCACTAG 480

Qy        522 CTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGATGAGGAGTAAGAACT 581
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGATGAGGAGTAAGAACT 540

Qy        582 CCCTTATCACTAACCCCCATC 602
                |||||||||||||||||||
Db        541 -CCTTATCACTAACCCCCATC 560


RESULT 8
AAX28676
ID   AAX28676 standard; cDNA; 501 BP.
XX
AC   AAX28676;
XX
DT   26-AUG-1999  (first entry)
XX
DE   Clone HP01738 encoding a human secretory signal protein (1).
XX
KW   Human; secretory signal protein sequence; cell membrane;
KW   proliferation; differentiation; carcinostatic agent; antigen;
KW   antibody; probe; hybridisation; gene therapy; HP01738; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..501
FT                   /*tag= a
FT                   /product= "secretory signal protein"
FT                   /note= "no termination codon given"
XX
PN   WO9918204-A2.
XX
PD   15-APR-1999.
XX
PF   05-OCT-1998;   98WO-JP04476.
XX
PR   08-OCT-1997;   97JP-0276268.
XX
PA   (PROT-) PROTEGENE INC.
PA   (SAGA ) SAGAMI CHEM RES CENT.
XX
PI   Kato S,  Kobayashi M,  Sekine S,  Yamaguchi T;
XX
DR   WPI; 1999-264020/22.
DR   P-PSDB; AAY03234.
XX
PT   Human proteins with secretory signal sequences and nucleotide
PT   sequences
XX
PS   Claim 3; Page 71; 84pp; English.
XX
CC   This is the nucleotide sequence of a clone encoding a human
CC   secretory signal protein sequence, used in the method of the
CC   invention. All of the proteins exist in the cell membrane, so are
CC   considered to be proteins controlling the proliferation and
CC   differentiation of the cells. They may be useful as carcinostatic
CC   agents or as antigens for preparing antibodies against the proteins.
CC   The cDNAs can be used as probes for gene diagnosis and gene sources
CC   for gene therapy, as well as for large-scale expression of the
CC   proteins.
XX
SQ   Sequence 501 BP; 101 A; 136 C; 139 G; 125 T; 0 other;

  Query Match             50.0%;  Score 501;  DB 20;  Length 501;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-110;
  Matches  501;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         34 ATGTTGACAGTCGCTCTCCTAGCCCTTCTCTGTGCCTCAGCCTCTGGCAATGCCATTCAG 93
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGTTGACAGTCGCTCTCCTAGCCCTTCTCTGTGCCTCAGCCTCTGGCAATGCCATTCAG 60

Qy         94 GCCAGGTCTTCCTCCTATAGTGGAGAGTATGGAAGTGGTGGTGGAAAGCGATTCTCTCAT 153
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCCAGGTCTTCCTCCTATAGTGGAGAGTATGGAAGTGGTGGTGGAAAGCGATTCTCTCAT 120

Qy        154 TCTGGCAACCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACACATACTAC 213
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCTGGCAACCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACACATACTAC 180

Qy        214 ATCGTAGGTCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGGTCGCAAC 273
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ATCGTAGGTCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGGTCGCAAC 240

Qy        274 GGAGACCTGGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTCTGGGAAG 333
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGAGACCTGGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTCTGGGAAG 300

Qy        334 TACAAGTGGTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCTGTCTTTT 393
```

```
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TACAAGTGGTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCTGTCTTTT 360

Qy        394 GGGAAAGACAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGTGCTCCGC 453
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGGAAAGACAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGTGCTCCGC 420

Qy        454 TTCATCAGTGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGATGTTTAC 513
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TTCATCAGTGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGATGTTTAC 480

Qy        514 CCCACTAGCTGCAGCAGATGC 534
                    |||||||||||||||||||||
Db        481 CCCACTAGCTGCAGCAGATGC 501


RESULT 9
AAK88295
ID   AAK88295 standard; cDNA; 661 BP.
XX
AC   AAK88295;
XX
DT   05-NOV-2001  (first entry)
XX
DE   Human digestive system antigen coding sequence SEQ ID NO: 611.
XX
KW   Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW   ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW   digestive system disorder; Meckel's diverticulum; ss.
XX
OS   Homo sapiens.
XX
PN   WO200155314-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01324.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
```

```
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-502630/55.
DR   P-PSDB; AAM92522.
XX
PT   Polynucleotides encoding digestive system antigens, useful for
PT   diagnosing, treating, preventing and/or prognozing disorders of the
PT   digestive system, particularly cancer and cancer metastases -
XX
PS   Claim 1; SEQ ID NO 611; 986pp; English.
XX
CC   The present invention provides the protein and coding sequences of a
CC   number of human digestive system antigens. These can be used in the
CC   diagnosis, treatment and prevention of digestive system disorders,
CC   including cancer, Meckel's diverticulum, bacterial or parasitic
CC   infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC   ulcerative colitis. The present sequence is a cDNA encoding a digestive
CC   system antigen of the invention.
XX
SQ   Sequence 661 BP; 146 A; 175 C; 180 G; 156 T; 4 other;

  Query Match             42.9%;  Score 429.4;  DB 22;  Length 661;
  Best Local Similarity   93.6%;  Pred. No. 2.8e-93;
  Matches  468;  Conservative    0;  Mismatches   29;  Indels    3;  Gaps    2;

Qy         27 CCCCAGAATGTTGACAGTCGCTCTCCTAGCCCTTCTCTGTGCCTCAGCCTCTGGCAATGC 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        159 CCCCAGAATGTTGACAGTCGCTCTCCTAGCCCTTCTCTGTGCCTCAGCCTCTGGCAATGC 218

Qy         87 CATTCAGGCCAGGTCTTCCTCCTATAGTGGAGAGTATGGAAGTGGTGGTGGAAAGCGATT 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        219 CATTCAGGCCAGGTCTTCCTCCTATAGTGGAGAGTATGGAAGTGGTGGTGGAAAGCGATT 278

Qy        147 CTCTCATTCTGGCAACCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACAC 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        279 CTCTCATTCTGGCAACCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACAC 338

Qy        207 ATACTACATCGTAGGTCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGG 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        339 ATACTACATCGTAGGTCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGG 398

Qy        267 TCGCAACGGAGACCTGGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTC 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        399 TCGCAACGGAGACCTGGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTC 458

Qy        327 TGGGAAGTACAAGTGGTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCT 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        459 TGGGAAGTACAAGTGGTACCTGAAGAAGCTGGTATTTGTGACAGAGAAGGGCCGCTATCT 518

Qy        387 GTCTTTTGGGAAAGACAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGT 446
              |||||||||||||||||||||||||||||||||||||||| |    ||| |   |
Db        519 GTCTTTTGGGAAAGACAGTGGCACAAGTTTCAATGCCGTCCCTTGCACCCCAACACCGTG 578

Qy        447 GCTCCGCTTCATCAGTGGCCGGTCTGGTTCTCTCA--TCGATGCCATTGGCCTGCAC-TG 503
                ||||||||||| ||| | | |||| |||||         |||||||| |||||| ||
Db        579 CTTCCGCTTCATCAATGGGCCGGCTGGGTCTCTTNATCGATGGCCATTGGNCTGCACTTG 638

Qy        504 GGATGTTTACCCCACTAGCT 523
              ||||||||||||||| |||||
Db        639 GGATGTTTACCCCANTAGCT 658


RESULT 10
AAF84674/c
ID   AAF84674 standard; DNA; 5086 BP.
XX
AC   AAF84674;
XX
DT   29-JUN-2001  (first entry)
XX
DE   Nucleotide sequence of plasmid pMTet1.
XX
KW   pMTet1; cytotoxic gene; selection; host cell isolation; ss.
XX
OS   Synthetic.
XX
PN   WO200125451-A2.
XX
PD   12-APR-2001.
XX
PF   22-SEP-2000; 2000WO-GB03655.
XX
PR   01-OCT-1999;   99US-0157072.
PR   25-JAN-2000; 2000GB-0001716.
PR   07-SEP-2000; 2000US-0656503.
XX
PA   (TRAN-) TRANSGENOMIC INC.
XX
PI   Hornby DPJ,  Matin MM;
XX
DR   WPI; 2001-273580/28.
XX
PT   One step cloning vector for quicker cloning of nucleic acid fragments
PT   comprises both a positive selection gene and a gene which protects
PT   against a selective agent -
XX
PS   Claim 23; Page 41-42; 49pp; English.
XX
CC   The present sequence represents plasmid pMTet1, which is used in the
CC   method of the invention. The specification describes one-step,
CC   non-plating methods for isolating transformed host cells which contain
CC   a vector incorporating a nucleic acid insert encoding a cytotoxic gene
CC   and a gene which provides the transformed host cell with a protective
CC   effect against the selection agent. The method uses a vector which
CC   comprises a cytotoxic gene permitting positive selection arranged so
CC   that it is insertionally inactivated by the incorporation into the vector
CC   of a heterologous nucleic acid insert; and a gene which provides a host
CC   cell transformed with the vector with a protective effect against the
CC   selection agent, but which does not significantly degrade the selection
CC   agent within a medium in which the host is present. The method allows
CC   the isolation of a transformed host cell which contains a vector
CC   incorporating a nucleic acid insert encoding a cytotoxic gene and a
CC   gene which provides the transformed host cell with a protective effect
CC   against the selection agent.
XX
```

```
SQ   Sequence 5086 BP; 1304 A; 1275 C; 1247 G; 1260 T; 0 other;

  Query Match            35.5%;  Score 355.8;  DB 22;  Length 5086;
  Best Local Similarity  80.2%;  Pred. No. 1.5e-75;
  Matches  390;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy        516 CACTAGCTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGATGAGGAGTA 575
              ||||| || ||  |  | | ||  ||    |   |  ||   || | |  |  | |  |
Db       3910 CACTAACTTCAATAATTTCAGATGCTTGCTTTTGAGTTAGATCCAATTTACTACGAATCA 3851

Qy        576 AGAACTCCCTTATCACTAACCCCCATCCAAATGGNNNNNNNNNNNNNNNNNNNNNNNNNN 635
              |  |    ||     |  |    |||||    |
Db       3850 ATTTCAATATTTCAGGTTTCATGGATCCATTGCGGCCGCACAGCAGCTTGGCGTAATCAT 3791

Qy        636 NNNNNNNNNNNNNTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAG 695
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3790 GGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAG 3731

Qy        696 CCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTG 755
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3730 CCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTG 3671

Qy        756 CGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAA 815
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3670 CGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAA 3611

Qy        816 TCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCA 875
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3610 TCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCA 3551

Qy        876 CTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGG 935
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3550 CTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGG 3491

Qy        936 TAATACGGTTATCCACAGAATCAGGGGGATACGCAGGAAAGAACATGTGAGCAAAAGGCC 995
              ||||||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db       3490 TAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCC 3431

Qy        996 AGCAAA 1001
              ||||||
Db       3430 AGCAAA 3425


RESULT 11
AAZ38921/c
ID   AAZ38921 standard; DNA; 5924 BP.
XX
AC   AAZ38921;
XX
DT   22-FEB-2000  (first entry)
XX
DE   hCAT1 binding human antibody molecule nucleotide sequence.
XX
KW   Virus-like particle; ligand; gene delivery; envelope protein; hCAT1;
KW   human cationic amino acid transporter 1; retrovirus; adenovirus;
KW   targeted delivery; gene therapy; balloon angioplasty; ss.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   EP960942-A2.
XX
PD   01-DEC-1999.
XX
PF   20-MAY-1999;   99EP-0201593.
XX
PR   20-MAY-1998;   98EP-0201693.
XX
PA   (INTR-) INTROGENE BV.
XX
PI   Van Es H,  Verlinden S,  Havenga M;
XX
DR   WPI; 2000-025491/03.
DR   P-PSDB; AAY56637.
XX
PT   New gene therapy vectors, useful for treating balloon angioplasty
PT   patients -
XX
PS   Example 3; Fig 16; 50pp; English.
XX
CC   The present invention describes a virus-like particle or gene delivery
CC   vehicle (I) provided with a ligand capable of binding to a human amino
CC   acid transporter. (I) is used to deliver genes to human cells or primate
CC   cells that express the hCAT1 amino acid transporter, such as endothelial,
CC   haematopoietic or smooth muscle cells, as part of a gene therapy regime.
CC   The vectors are especially useful for providing local applications of
CC   adenoviral vector to patients with restenosis following balloon
CC   angioplasty, where smooth muscle cells need to be transduced with ceNOS
CC   cDNA, for example. (I) may also be used to pseudotype recombinant type C
CC   retrovirus including murine leukemia retroviruses and lentiviruses. In
CC   addition (I) may be used in functional genomics where transduction of as
CC   many cell types as possible is required. The new gene delivery vehicles
CC   transduce DNA more efficiently and specifically into tissues that are
CC   hard to transform, such as endothelial cells or smooth muscle cells as
CC   compared to a wildtype adenoviral vector. This increased specificity
CC   results in lower multiplicities of infection which can occur with prior
CC   art vectors, so preventing tissue toxicity. In addition the new vectors
CC   allow DNA to be transduced into cells that are in low abundance in cell
CC   mixtures and tissues, which increases their efficiency for use as gene
CC   therapy vehicles. The alteration of the ligand on the viral envelope
CC   increases the potential host cell range of these vehicles. The present
CC   sequence encodes a hCAT1 binding human antibody molecule from the
CC   present invention.
XX
SQ   Sequence 5924 BP; 1454 A; 1487 C; 1517 G; 1466 T; 0 other;

  Query Match            35.2%;  Score 352.4;  DB 21;  Length 5924;
  Best Local Similarity  88.8%;  Pred. No. 1e-74;
  Matches  356;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy        601 TCCAAATGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCCTGTGTGAAA 660
              ||||||                                            ||||||||||||
Db       2257 TCCAAAAAAAAAGGCTCCAAAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAA 2198

Qy        661 TTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2197 TTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTG 2138

Qy        721 GGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2137 GGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCA 2078

Qy        781 GTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2077 GTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGG 2018

Qy        841 TTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2017 TTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCG 1958

Qy        901 GCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1957 GCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGG 1898

Qy        961 GGGATACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1001
              ||   ||||||||||||||||||||||||||||||||||||
Db       1897 GGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1857


RESULT 12
AAZ38770/c
ID   AAZ38770 standard; DNA; 5925 BP.
XX
AC   AAZ38770;
XX
DT   08-FEB-2000  (first entry)
```

```
XX
DE   HCAT1 clone 25 antibody nucleotide sequence SEQ ID NO:57.
XX
KW   Human cationic amino acid transporter; hCAT1; targeted delivery;
KW   gene delivery; virus-like particle; retroviral envelope molecule;
KW   infection; gene therapy; restenosis; balloon angioplasty;
KW   smooth muscle cell; transduction; ss.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   EP959136-A1.
XX
PD   24-NOV-1999.
XX
PF   20-MAY-1998;   98EP-0201693.
XX
PR   20-MAY-1998;   98EP-0201693.
XX
PA   (INTR-) INTROGENE BV.
XX
DR   WPI; 2000-001283/01.
DR   P-PSDB; AAY56286, AAY56287.
XX
PT   New virus-like particle or gene delivery vehicle, useful for gene
PT   therapy -
XX
PS   Claim 13; Fig 16; 66pp; English.
XX
CC   The present invention describes a virus-like particle or gene delivery
CC   vehicle comprising a ligand capable of binding to a human amino acid
CC   transporter. The method is useful for the target delivery of substances
CC   to cells e.g. gene therapy. A human cationic amino acid transporter
CC   (hCAT1) targeted adenovirus is useful for local applications of
CC   adenoviral vector e.g. in patients with restenosis following balloon
CC   angioplasty where smooth muscle cells need to be transduced with an
CC   adenoviral vector carrying the CeNOS cDNA. More efficient transduction
CC   of tissues can be carried out therefore resulting in lower
CC   multiplicity's of infections that can be used and therefore less vector
CC   associated toxicity to the tissues surrounding the target cells. AAZ38737
CC   to AAZ38770, and AAY56264 to AAY56287 represent sequences used in the
CC   exemplification of the present invention.
XX
SQ   Sequence 5925 BP; 1454 A; 1487 C; 1517 G; 1467 T; 0 other;

  Query Match            35.2%;  Score 352.4;  DB 21;  Length 5925;
  Best Local Similarity  88.8%;  Pred. No. 1e-74;
  Matches  356;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy        601 TCCAAATGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCCTGTGTGAAA 660
              ||||||                                             ||||||||||||
Db       2258 TCCAAAAAAAAGGCTCCAAAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAA 2199

Qy        661 TTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2198 TTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTG 2139

Qy        721 GGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2138 GGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCA 2079

Qy        781 GTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2078 GTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGG 2019

Qy        841 TTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2018 TTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCG 1959

Qy        901 GCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1958 GCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGG 1899

Qy        961 GGGATACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1001
              ||   ||||||||||||||||||||||||||||||||||||
Db       1898 GGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1858


RESULT 13
AAZ39782
ID   AAZ39782 standard; DNA; 3789 BP.
XX
AC   AAZ39782;
XX
DT   06-MAR-2000  (first entry)
XX
DE   Plasmid pGB1MMN nucleotide sequence.
XX
KW   Manganese superoxide dismutase; MnSOD; ECSOD; proteoglycan; human;
KW   endothelial; oxidative damage; pulmonary inflammatory injury; cancer;
KW   lung disease; hypoxia; ischemia; reperfusion injury; arthritis;
KW   hyperoxia; atherosclerosis; lupus erythematosis; hypertension;
KW   neutrophil-mediated inflammation; mutant; ss.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9958547-A1.
XX
PD   18-NOV-1999.
XX
PF   06-MAY-1999;   99WO-US09921.
XX
PR   08-MAY-1998;   98US-0075019.
XX
PA   (WEBB-) WEBB-WARING INST BIOMEDICAL RES.
XX
PI   McCord JM,  Gao B,  Flores SC;
XX
DR   WPI; 2000-062283/05.
DR   P-PSDB; AAY55849.
XX
PT   Modified manganese superoxide dismutase, methods of production and
PT   antibodies  -
XX
PS   Claim 11; Page 77-79; 83pp; English.
XX
CC   The invention provides a nucleic acid molecule encoding a genetically
CC   modified manganese superoxide dismutase (MnSOD). The nucleic acid
CC   comprises: a first nucleic acid sequence encoding an enzymatically active
CC   portion of MnSOD; and a second nucleic acid sequence encoding a peptide
CC   (ECSOD) that binds to polyanionic polysaccharides or proteoglycans on
CC   endothelial cell surfaces. The protein protects a mammal, especially
CC   humans, from oxidative damage and especially from conditions consisting
CC   of pulmonary inflammatory injury, lung disease, cancer, hypoxia, ischemia
CC   reperfusion injury, hyperoxia, atherosclerosis, arthritis, lupus
CC   erythematosis, hypertension and neutrophil-mediated inflammation. The
CC   lung disease is infant or adult respiratory distress syndrome,
CC   interstitial lung disease or asthma. The mutant MnSOD also protects
CC   organs of mammals from pre- and post-transplantation oxidative damage.
CC   MnSOD is positively charged at physiological pH and has a longer plasma
CC   half-life of about 4 hours compared to 6-15 minutes for cytosolic Cu,Zn
CC   SOD (following intravenous injection). ECSOD has a substantial advantage
CC   over both Cu, ZnSOD and MnSOD because of its ability to bind to the
CC   endothelium. ECSOD is hard to produce in mammalian cell culture systems,
CC   so a fusion of ECSOD and MnSOD therefore overcomes these problems. The
CC   present sequence represents the nucleotide sequence of the plasmid
CC   pGB1MMN. This contains the coding sequence of the human MnSOD mutant
CC   protein which comprises the MnSOD sequence, and sa 78 basepair ECSOD
CC   tail-encoding fragment.
XX
SQ   Sequence 3789 BP; 991 A; 933 C; 952 G; 913 T; 0 other;

  Query Match            35.2%;  Score 352.2;  DB 21;  Length 3789;
  Best Local Similarity  88.4%;  Pred. No. 1.1e-74;
  Matches  357;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;
```

```
Qy        598 CCATCCAAATGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCCTGTGTG 657
              || || |  ||                                              |||||||||
Db       1529 CCGTCGACCTGCAGGGCATGCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTG 1588

Qy        658 AAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGC 717
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1589 AAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGC 1648

Qy        718 CTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTT 777
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1649 CTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTT 1708

Qy        778 CCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGG 837
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1709 CCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGG 1768

Qy        838 CGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGT 897
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1769 CGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGT 1828

Qy        898 TCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATC 957
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1829 TCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATC 1888

Qy        958 AGGGGGATACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1001
              |||||   |||||||||||||||||||||||||||||||||||
Db       1889 AGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1932


RESULT 14
AAQ34611
ID   AAQ34611 standard; DNA; 5393 BP.
XX
AC   AAQ34611;
XX
DT   13-MAY-1993  (first entry)
XX
DE   Plasmid pR3.
XX
KW   Plasmid; pR3; Eimeria; merozoite; surface; antigen; precusor;
KW   protein; recombinant vector; microbial host; parasite; vaccine;
KW   coccidiosis; poultry; ss.
XX
OS   Synthetic.
XX
PN   EP522482-A.
XX
PD   13-JAN-1993.
XX
PF   06-JUL-1992;   92EP-0111407.
XX
PR   12-JUL-1991;   91US-0729099.
XX
PA   (HOFF ) HOFFMANN LA ROCHE & CO AG F.
XX
PI   Binger M,  Pasamontes L;
XX
DR   WPI; 1993-010336/02.
DR   P-PSDB; AAR30993.
XX
PT   Immunogenic specified polypeptide and fragments - capable of
PT   inducing an immune response against eimeria parasites and free
PT   from proteins produced by these parasites
XX
PS   Disclosure; Fig 14; 72pp; English.
XX
CC   This sequence represents the plasmid pR3.  This plasmid contains the
CC   coding sequence for the E. merozoite surface antigen precusor protein.
CC   This recombinant vector to was used to transform a microbial host to
CC   produce the precusor protein.  The protein encoded by this sequence
CC   is the antigen of Eimeria parasites and may be used in a vaccine
CC   against coccidiosis in poultry.
XX
SQ   Sequence 5393 BP; 1541 A; 1175 C; 1266 G; 1411 T; 0 other;

  Query Match            35.2%;  Score 352.2;  DB 14;  Length 5393;
  Best Local Similarity  88.4%;  Pred. No. 1.1e-74;
  Matches  357;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        598 CCATCCAAATGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCCTGTGTG 657
              |||| |  | ||                                              |||||||||
Db       3133 CCACCGTAGTGTATATAGGATCTGCTCCCCGTAATCATGGTCATAGCTGTTTCCTGTGTG 3192

Qy        658 AAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGC 717
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3193 AAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGC 3252

Qy        718 CTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTT 777
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3253 CTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTT 3312

Qy        778 CCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGG 837
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3313 CCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGG 3372

Qy        838 CGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGT 897
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3373 CGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGT 3432

Qy        898 TCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATC 957
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3433 TCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATC 3492

Qy        958 AGGGGGATACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1001
              |||||   |||||||||||||||||||||||||||||||||||
Db       3493 AGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 3536


RESULT 15
AAD10265
ID   AAD10265 standard; DNA; 3983 BP.
XX
AC   AAD10265;
XX
DT   24-SEP-2001  (first entry)
XX
DE   pUC/Ac vector DNA.
XX
KW   GAT-ZM1; seed purity; food; feed; transgenic plant identification;
KW   quality determination; phosphinothricin-acetyl-transferase; pat gene;
KW   pUC/Ac vector; beta-lactamase gene; herbicide-resistant corn; cyclic;
KW   circular; ds.
XX
OS   Chimeric - Cauliflower mosaic virus.
OS   Chimeric - Streptomyces viridochromogenes.
OS   Chimeric - Synthetic.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    412..618
FT                   /*tag= a
FT                   /note= "35S terminator from Cauliflower mosaic virus
FT                   from pDH51 vector"
FT   primer_bind     507..527
FT                   /*tag= b
FT                   /bound_moiety= "MDB420 primer"
FT   primer_bind     606..626
FT                   /*tag= c
FT                   /bound_moiety= "MDB411 primer"
FT   misc_feature    619..636
FT                   /*tag= d
FT                   /note= "Synthetic polylinker"
FT   misc_feature    637..1188
FT                   /*tag= e
FT                   /note= "Streptomyces viridochromogenes pat gene"
FT   primer_bind     715..736
```

```
FT                   /*tag= f
FT                   /bound_moiety= "MDB391 primer"
FT   misc_feature    1189..1216
FT                   /*tag= g
FT                   /note= "Synthetic polylinker"
FT   misc_feature    1217..1746
FT                   /*tag= h
FT                   /note= "35S promoter from Cauliflower mosaic virus
FT                   from pDH51 vector"
FT   primer_bind     complement (1326..1345)
FT                   /*tag= i
FT                   /bound_moiety= "MB522 primer"
FT   misc_feature    1747..411
FT                   /*tag= j
FT                   /note= "Sequence of the vector pUC18 including the beta
FT                   lactamase gene and the origin of the replication fork"
FT   misc_signal     2164
FT                   /*tag= k
FT                   /note= "Origin of replication fork"
FT   misc_feature    2923..3783
FT                   /*tag= l
FT                   /note= "Beta-lactamase gene"
FT   primer_bind     complement (3401..3422)
FT                   /*tag= m
FT                   /bound_moiety= "MB522 primer"
FT   primer_bind     complement (3435..3457)
FT                   /*tag= n
FT                   /bound_moiety= "MB522 primer"
XX
PN   WO200151654-A2.
XX
PD   19-JUL-2001.
XX
PF   09-JAN-2001; 2001WO-EP00306.
XX
PR   11-JAN-2000; 2000US-0481049.
XX
PA   (AVET ) AVENTIS CROPSCIENCE NV.
XX
PI   De Beuckeleer M;
XX
DR   WPI; 2001-451864/48.
XX
PT   Identifying elite event GAT-ZM1 in biological samples for confirming
PT   seed purify in seed samples, by detecting GAT-ZM1 specific region with
PT   specific primer/probe which specifically recognizes its flanking
PT   regions -
XX
PS   Example 1; Page 29-30; 34pp; English.
XX
CC   The patent discloses methods and kits for identifying elite event
CC   GAT-ZM1 in biological samples. The method involves detecting a GAT-ZM1
CC   specific region with a specific primer or probe which specifically
CC   recognises the 5' or 3' flanking region of GAT-ZM1. This method is
CC   useful for confirming seed purity in seed samples, and for screening
CC   seeds for the presence of GAT-ZM1 in samples of seed lots. It is useful
CC   for identifying GAT-ZM1 in plants, plant material or in products such
CC   as food or feed products comprising or derived from plant material, to
CC   identify transgenic plant material for purposes of segregation between
CC   transgenic and non-transgenic material, and to determine the quality
CC   of plant material comprising GAT-ZM1. It allows rapid and unequivocal
CC   identification of elite event GAT-ZM1 in biological samples. The
CC   invention also relates to 5' and/or 3' flanking regions of GAT-ZM1 as
CC   well as to the specific primers and probes developed from the 5' and/or
CC   3' flanking sequences of GAT-ZM1. The present sequence is pUC/Ac vector
CC   which comprises the coding sequence of a pat gene encoding the enzyme
CC   phosphinothricin-acetyl-transferase, under the control of the
CC   constitutive 35S promoter from Cauliflower Mosaic Virus (CaMV). It also
CC   contains a 35S terminator from CaMV, synthetic polylinker sequences and
CC   beta-lactamase gene. This vector is used to develop herbicide-resistant
CC   corn.
XX
SQ   Sequence 3983 BP; 1002 A; 991 C; 958 G; 1032 T; 0 other;
```

```
  Query Match             35.2%;  Score 352;  DB 22;  Length 3983;
  Best Local Similarity   88.0%;  Pred. No. 1.2e-74;
  Matches  358;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy        595 CCCCCATCCAAATGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCCTGT 654
              || | ||   |||                                                  ||||||
Db       1720 CCTCTATTTGAATCTTTGACTCCATGGGAATTCGTAATCATGGTCATAGCTGTTTCCTGT 1779

Qy        655 GTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAA 714
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1780 GTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAA 1839

Qy        715 AGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGC 774
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1840 AGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGC 1899

Qy        775 TTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAG 834
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1900 TTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAG 1959

Qy        835 AGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGT 894
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1960 AGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGT 2019

Qy        895 CGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGA 954
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2020 CGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGA 2079

Qy        955 ATCAGGGGGATACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1001
              ||||||||  ||||||||||||||||||||||||||||||||||||
Db       2080 ATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 2126


Search completed: July 16, 2003, 11:09:54
Job time : 175.989 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 16, 2003, 08:51:42 ; Search time 1092.78 Seconds
(without alignments)
14835.236 Million cell updates/sec

Title: US-09-981-353-24
Perfect score: 1001
Sequence: 1 atgattacgaattcgagctc.....gtgagcaaaaggccagcaaa 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	575	57.4	605	9	AI732905	AI732905 of72a12.x
c	2	544.4	54.4	581	13	BM021151	BM021151 ie76h03.x
c	3	542.8	54.2	581	13	BM023112	BM023112 ie79e10.x
c	4	542.4	54.2	602	9	AI833168	AI833168 at75h09.x
	5	530	52.9	549	14	BM934914	BM934914 UI-M-BH3-
	6	521.4	52.1	633	13	BI763402	BI763402 603047590
	7	517.8	51.7	572	13	BM021411	BM021411 ie76h03.y
c	8	497.8	49.7	566	9	AI991154	AI991154 wu38h12.x
c	9	469.4	46.9	520	14	BQ028047	BQ028047 UI-H-CO0-
c	10	447	44.7	490	14	BM987195	BM987195 UI-H-CO0-
c	11	441	44.1	485	14	BM987408	BM987408 UI-H-CO0-
c	12	435.4	43.5	492	14	BM987303	BM987303 UI-H-CO0-
c	13	435.2	43.5	497	9	AA828206	AA828206 of05d11.s
c	14	426.4	42.6	455	9	AI749030	AI749030 at38b05.x
c	15	423.2	42.3	467	10	AW001526	AW001526 wu34a06.x
	16	419.6	41.9	489	13	BM023370	BM023370 ie79e10.y
c	17	418.4	41.8	448	9	AI984533	AI984533 wu22h06.x
c	18	416.8	41.6	449	9	AI880265	AI880265 at47d07.x
c	19	410.4	41.0	456	14	BM987839	BM987839 UI-H-CO0-
c	20	409.8	40.9	496	10	AW000842	AW000842 wu46c02.x
c	21	400.4	40.0	466	10	AW001333	AW001333 wu28h11.x
c	22	394.4	39.4	464	9	AA937899	AA937899 of72a12.s
c	23	393	39.3	446	9	AI984522	AI984522 wu22g02.x
c	24	386.2	38.6	440	10	AW456570	AW456570 UI-M-BH3-
c	25	382.4	38.2	427	10	AW001511	AW001511 wu33g01.x
	26	375.2	37.5	467	10	AW844215	AW844215 RC4-CN004
c	27	374	37.4	418	14	BM987670	BM987670 UI-H-CO0-
	28	370.4	37.0	683	12	BF585282	BF585282 602101945
	29	370.4	37.0	743	14	BQ960080	BQ960080 AGENCOURT
	30	370.2	37.0	550	12	BG879223	BG879223 ib60d09.y
	31	370.2	37.0	564	12	BG311182	BG311182 ib21a09.y
	32	370.2	37.0	647	13	BG967068	BG967068 602833913
	33	370.2	37.0	662	14	BQ952658	BQ952658 AGENCOURT
	34	370.2	37.0	666	12	BF583056	BF583056 602097944
	35	370.2	37.0	666	13	BG967428	BG967428 602833573
	36	370.2	37.0	666	14	BQ954539	BQ954539 AGENCOURT
	37	370.2	37.0	668	14	BQ921513	BQ921513 AGENCOURT
	38	370.2	37.0	674	12	BF577589	BF577589 602092454
	39	370.2	37.0	675	14	BQ961186	BQ961186 AGENCOURT
	40	370.2	37.0	676	14	BQ947379	BQ947379 AGENCOURT
	41	370.2	37.0	677	14	BQ929155	BQ929155 AGENCOURT
	42	370.2	37.0	679	12	BF579244	BF579244 602093414
	43	370.2	37.0	686	13	BG966346	BG966346 602832832
	44	370.2	37.0	689	14	BQ939500	BQ939500 AGENCOURT
	45	370.2	37.0	689	14	BQ945842	BQ945842 AGENCOURT

ALIGNMENTS

RESULT 1
AI732905/c
LOCUS AI732905 605 bp mRNA linear EST 13-DEC-1999
DEFINITION of72a12.x5 NCI_CGAP_Co8 Homo sapiens cDNA clone IMAGE:1435870 3' similar to TR:O60844 O60844 HOMOLOG OF RAT ZYMOGEN GRANULE MEMBRANE PROTEIN. ;, mRNA sequence.
ACCESSION AI732905
VERSION AI732905.1 GI:5054018
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 605)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

```
          This read is a RESEQUENCE of a previously sequenced human clone
          Original clone citation: National Cancer Institute, Cancer Genome
          Anatomy Project (CGAP), Tumor Gene Index
          This read has been verified (found to hit its original self in the
          correct orientation)
          Insert Length: 725   Std Error: 0.00
          Seq primer: -40UP from Gibco
          High quality sequence stop: 469.
FEATURES             Location/Qualifiers
     source          1. .605
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1435870"
                     /clone_lib="NCI_CGAP_Co8"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B"
                     /note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; 1st strand cDNA was prepared from
                     colon adenocarcinoma, and was then primed with a Not I -
                     oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                     RI adaptors (Pharmacia), digested with Not I and cloned
                     into the Not I and Eco RI sites of the modified pT7T3
                     vector.  Library is normalized. Library was constructed by
                     Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      148 a    160 c    160 g    136 t      1 others
ORIGIN

  Query Match             57.4%;  Score 575;  DB 9;  Length 605;
  Best Local Similarity   99.8%;  Pred. No. 1.5e-154;
  Matches  575;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         34 ATGTTGACAGTCGCTCTCCTAGCCCTTCTCTGTGCCTCAGCCTCTGGCAATGCCATTCAG 93
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        605 ATGTTGACAGTCGCTCTCCTAGCCCTTCTCTGTGCCTCAGCCTCTGGCAATGCCATTCAG 546

Qy         94 GCCAGGTCTTCCTCCTATAGTGGAGAGTATGGAAGTGGTGGTGGAAAGCGATTCTCTCAT 153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        545 GCCAGGTCTTCCTCCTATAGTGGAGAGTATGGAAGTGGTGGTGGAAAGCGATTCTCTCAT 486

Qy        154 TCTGGCAACCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACACATACTAC 213
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        485 TCTGGCAACCAGTTGGACGGCNCCATCACCGCCCTCCGGGTCCGAGTCAACACATACTAC 426

Qy        214 ATCGTAGGTCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGGTCGCAAC 273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        425 ATCGTAGGTCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGGTCGCAAC 366

Qy        274 GGAGACCTGGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTCTGGGAAG 333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        365 GGAGACCTGGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTCTGGGAAG 306

Qy        334 TACAAGTGGTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCTGTCTTTT 393
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 TACAAGTGGTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCTGTCTTTT 246

Qy        394 GGGAAAGACAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGTGCTCCGC 453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        245 GGGAAAGACAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGTGCTCCGC 186

Qy        454 TTCATCAGTGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGATGTTTAC 513
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        185 TTCATCAGTGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGATGTTTAC 126

Qy        514 CCCACTAGCTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGATGAGGAG 573
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        125 CCCACTAGCTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGATGAGGAG 66

Qy        574 TAAGAACTCCCTTATCACTAACCCCCATCCAAATGG 609
              ||||||||||||||||||||||||||||||||||||
Db         65 TAAGAACTCCCTTATCACTAACCCCCATCCAAATGG 30


RESULT 2
```

```
BM021151/c
LOCUS       BM021151                581 bp    mRNA    linear   EST 12-MAR-2002
DEFINITION  ie76h03.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
            cDNA clone IMAGE:5673053 3' similar to TR:O60844 O60844 HOMOLOG OF
            RAT ZYMOGEN GRANULE MEMBRANE PROTEIN. ;, mRNA sequence.
ACCESSION   BM021151
VERSION     BM021151.1  GI:16535507
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 581)
  AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
            ,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.
            , Jackson,Y. and Bowers,Y.
  TITLE     Endocrine Pancreas Consortium
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Douglas Melton DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Juliana Brown
            (brown@fas.harvard.edu) This sequence now available from the IMAGE
            consortium, for clone orders contact: info@image.llnl.gov
            High quality sequence stop: 447.
FEATURES             Location/Qualifiers
     source          1. .581
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5673053"
                     /clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
                     /sex="Both"
                     /tissue_type="Islets of Langerhans"
                     /dev_stage="Adult"
                     /lab_host="DH10B"
                     /note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;
                     Site_2: Sal 1; Starting library constructed using
                     SuperScript Plasmid Library kit  (Life Technologies). cDNA
                     made by oligo-dT priming. Size-selected by  column
                     fractionation; average insert size 1.08 kb. Library was
                     amplified  once on solid support and plasmid DNA from
                     library was prepared. The  library DNA was normalized by
                     method #4 from Bonaldo, Lennon, and Soares  1996 Genome
                     Research 6:791-806; 0.5 microgram single-stranded library
                     plasmid DNA was mixed with 5 micrograms PCR product
                     representing library  inserts and hybridized to an Ecot of
                     20. Single-stranded (unhybridized)  plasmids were isolated
                     by hydroxyapatite chromatography and used to make  this
                     library."
BASE COUNT      139 a    155 c    150 g    137 t
ORIGIN

  Query Match             54.4%;  Score 544.4;  DB 13;  Length 581;
  Best Local Similarity   99.8%;  Pred. No. 9.9e-146;
  Matches  545;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         64 TGTGCCTCAGCCTCTGGCAATGCCATTCAGGCCAGGTCTTCCTCCTATAGTGGAGAGTAT 123
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 TGTGCCTCAGCCTCTGGCAATGCCATTCAGGCCAGGTCTTCCTCCTATAGTGGAGAGTAT 522

Qy        124 GGAAGTGGTGGTGGAAAGCGATTCTCTCATTCTGGCAACCAGTTGGACGGCCCCATCACC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 GGAAGTGGTGGTGGAAAGCGATTCTCTCATTCTGGCAACCAGTTGGACGGCCCCATCACC 462
```

```
Qy        184 GCCCTCCGGGTCCGAGTCAACACATACTACATCGTAGGTCTTCAGGTGCGCTATGGCAAG 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 GCCCTCCGGGTCCGAGTCAACACATACTACATCGTAGGTCTTCAGGTGCGCTATGGCAAG 402

Qy        244 GTGTGGAGCGACTATGTGGGTGGTCGCAACGGAGACCTGGAGGAGATCTTTCTGCACCCT 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 GTGTGGAGCGACTATGTGGGTGGTCGCAACGGAGACCTGGAGGAGATCTTTCTGCACCCT 342

Qy        304 GGGGAATCAGTGATCCAGGTTTCTGGGAAGTACAAGTGGTACCTGAAGAAGCTGGTATTT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 GGGGAATCAGTGATCCAGGTTTCTGGGAAGTACAAGTGGTACCTGAAGAAGCTGGTATTT 282

Qy        364 GTGACAGACAAGGGCCGCTATCTGTCTTTTGGGAAAGACAGTGGCACAAGTTTCAATGCC 423
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 GTGACAGACAAGGGCCGCTATCTGTCTTTTGGGAAAGACAGTGGCACAAGTTTCAATGCC 222

Qy        424 GTCCCCTTGCACCCCAACACCGTGCTCCGCTTCATCAGTGGCCGGTCTGGTTCTCTCATC 483
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        221 GTCCCCTTGCACACCCAACACCGTGCTCCGCTTCATCAGTGGCCGGTCTGGTTCTCTCATC 162

Qy        484 GATGCCATTGGCCTGCACTGGGATGTTTACCCCACTAGCTGCAGCAGATGCTGAGCCTCC 543
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GATGCCATTGGCCTGCACTGGGATGTTTACCCCACTAGCTGCAGCAGATGCTGAGCCTCC 102

Qy        544 TCTCCTTGGCAGGGGCACTGTGATGAGGAGTAAGAACTCCCTTATCACTAACCCCCATCC 603
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 TCTCCTTGGCAGGGGCACTGTGATGAGGAGTAAGAACTCCCTTATCACTAACCCCCATCC 42

Qy        604 AAATGG 609
              ||||||
Db         41 AAATGG 36
```

RESULT 3
BM023112/c
LOCUS BM023112 581 bp mRNA linear EST 12-MAR-2002
DEFINITION ie79e10.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens cDNA clone IMAGE:5673306 3' similar to TR:O60844 O60844 HOMOLOG OF RAT ZYMOGEN GRANULE MEMBRANE PROTEIN. ;, mRNA sequence.
ACCESSION BM023112
VERSION BM023112.1 GI:16537468
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 581)
AUTHORS Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K., Lemishka,I., Sceance,M., Brestelli,J., Gradwohl,G., Clifton,S., Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A., Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas ,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T. , Jackson,Y. and Bowers,Y.
TITLE Endocrine Pancreas Consortium
JOURNAL Unpublished (2000)
COMMENT Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Juliana Brown (brown@fas.harvard.edu) This sequence now available from the IMAGE consortium, for clone orders contact: info@image.llnl.gov
High quality sequence stop: 344.
FEATURES Location/Qualifiers
source 1. .581
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5673306"
/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
/sex="Both"
/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="DH10B"
/note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1; Site_2: Sal 1; Starting library constructed using SuperScript Plasmid Library kit (Life Technologies). cDNA made by oligo-dT priming. Size-selected by column fractionation; average insert size 1.08 kb. Library was amplified once on solid support and plasmid DNA from library was prepared. The library DNA was normalized by method #4 from Bonaldo, Lennon, and Soares 1996 Genome Research 6:791-806; 0.5 microgram single-stranded library plasmid DNA was mixed with 5 micrograms PCR product representing library inserts and hybridized to an Ecot of 20. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."
BASE COUNT 140 a 155 c 149 g 137 t
ORIGIN

Query Match 54.2%; Score 542.8; DB 13; Length 581;
Best Local Similarity 99.6%; Pred. No. 2.8e-145;
Matches 544; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy         64 TGTGCCTCAGCCTCTGGCAATGCCATTCAGGCCAGGTCTTCCTCCTATAGTGGAGAGTAT 123
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 TGTGCTTCAGCCTCTGGCAATGCCATTCAGGCCAGGTCTTCCTCCTATAGTGGAGAGTAT 522

Qy        124 GGAAGTGGTGGTGGAAAGCGATTCTCTCATTCTGGCAACCAGTTGGACGGCCCCATCACC 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 GGAAGTGGTGGTGGAAAGCGATTCTCTCATTCTGGCAACCAGTTGGACGGCCCCATCACC 462

Qy        184 GCCCTCCGGGTCCGAGTCAACACATACTACATCGTAGGTCTTCAGGTGCGCTATGGCAAG 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 GCCCTCCGGGTCCGAGTCAACACATACTACATCGTAGGTCTTCAGGTGCGCTATGGCAAG 402

Qy        244 GTGTGGAGCGACTATGTGGGTGGTCGCAACGGAGACCTGGAGGAGATCTTTCTGCACCCT 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 GTGTGGAGCGACTATGTGGGTGGTCGCAACGGAGACCTGGAGGAGATCTTTCTGCACCCT 342

Qy        304 GGGGAATCAGTGATCCAGGTTTCTGGGAAGTACAAGTGGTACCTGAAGAAGCTGGTATTT 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 GGGGAATCAGTGATCCAGGTTTCTGGGAAGTACAAGTGGTACCTGAAGAAGCTGGTATTT 282

Qy        364 GTGACAGACAAGGGCCGCTATCTGTCTTTTGGGAAAGACAGTGGCACAAGTTTCAATGCC 423
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 GTGACAGACAAGGGCCGCTATCTGTCTTTTGGGAAAGACAGTGGCACAAGTTTCAATGCC 222

Qy        424 GTCCCCTTGCACCCCAACACCGTGCTCCGCTTCATCAGTGGCCGGTCTGGTTCTCTCATC 483
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 GTCCCCTTGAACCCCAACACCGTGCTCCGCTTCATCAGTGGCCGGTCTGGTTCTCTCATC 162

Qy        484 GATGCCATTGGCCTGCACTGGGATGTTTACCCCACTAGCTGCAGCAGATGCTGAGCCTCC 543
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GATGCCATTGGCCTGCACTGGGATGTTTACCCCACTAGCTGCAGCAGATGCTGAGCCTCC 102

Qy        544 TCTCCTTGGCAGGGGCACTGTGATGAGGAGTAAGAACTCCCTTATCACTAACCCCCATCC 603
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 TCTCCTTGGCAGGGGCACTGTGATGAGGAGTAAGAACTCCCTTATCACTAACCCCCATCC 42

Qy        604 AAATGG 609
              ||||||
Db         41 AAATGG 36
```

RESULT 4
AI833168/c
LOCUS AI833168 602 bp mRNA linear EST 13-JUL-1999
DEFINITION at75h09.x1 Barstead colon HPLRB7 Homo sapiens cDNA clone

```
            IMAGE:2377889 3' similar to TR:O60844 O60844 HOMOLOG OF RAT ZYMOGEN
            GRANULE MEMBRANE PROTEIN. ;, mRNA sequence.
ACCESSION   AI833168
VERSION     AI833168.1  GI:5455148
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 602)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
            ,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Possible reversed clone: similarity on wrong strand
            Seq primer: -40UP from Gibco
            High quality sequence stop: 214.
FEATURES             Location/Qualifiers
     source          1. .602
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2377889"
                     /clone_lib="Barstead colon HPLRB7"
                     /sex="male"
                     /dev_stage="adult, age 25"
                     /lab_host="DH10B (phage resistant)"
                     /note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     [5' AATTCACTAGTAAT 3' and 5' ATTACTAGTG 3'], digested
                     with Not I and cloned into the Not I and Eco RI sites of
                     the modified pT7T3 vector. Library constructed by Bob
                     Barstead."
BASE COUNT      152 a    159 c    161 g    129 t      1 others
ORIGIN

  Query Match           54.2%;  Score 542.4;  DB 9;  Length 602;
  Best Local Similarity  98.4%;  Pred. No. 3.8e-145;
  Matches  568;  Conservative    0;  Mismatches    7;  Indels    2;  Gaps    2;

Qy         34 ATGTTGACAGTCGCTCTCCTAGCCCTTCTCTGTGCCTCAGCCTCTGGCAATGCCATTCAG 93
              |||||||||||||||||||||||||||||||| ||||||||||||||| |||||||
Db        602 ATGTTGACAGTCGCTCTCCTAGCCCTTCTCTGTGCNTCAGCCTCTGGCAATG-CATTCAG 544

Qy         94 GCCAGGTCTTCCTCCTATAGTGGAGAGTATGGAAGTGGTGGTGGAAAGCGATTCTCTCAT 153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        543 GCCAGGTCTTCCTCCTATAGTGGAGAGTATGGAAGTGGTGGTGGAAAGCGATTCTCTCAT 484

Qy        154 TCTGGCAACCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACACATACTAC 213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        483 TCTGGCAACCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACACATACTAC 424

Qy        214 ATCGTAGGTCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGGTCGCAAC 273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 ATCGTAGGTCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGGTCGCAAC 364

Qy        274 GGAGACCTGGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTCTGGGAAG 333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 GGAGACCTGGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTCTGGGAAG 304

Qy        334 TACAAGTGGTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGG-CCGCTATCTGTCTTT 392
              |||||||||||||||||||||||||||||||||||||||||||| || || | |||||||
Db        303 TACAAGTGGTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCCCCTTTTTGTCTTT 244

Qy        393 TGGGAAAGACAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGTGCTCCG 452
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 TGGGAAAGACAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGTGCTCCG 184

Qy        453 CTTCATCAGTGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGATGTTTA 512
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        183 CTTCATCAGTGGCCGGTTTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGATGTTTA 124

Qy        513 CCCCACTAGCTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGATGAGGA 572
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        123 CCCCACTAGCTGCAGCAGATGTTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGATGAGGA 64

Qy        573 GTAAGAACTCCCTTATCACTAACCCCCATCCAAATGG 609
              |||||||||||||||||| ||||||||||||||||||
Db         63 GTAAGAACTCCCTTATCCCTAACCCCCATCCAAATGG 27


RESULT 5
BM934914
LOCUS       BM934914                549 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  UI-M-BH3-aqs-e-10-0-UI.r1 NIH_BMAP_M_S4 Mus musculus cDNA clone
            UI-M-BH3-aqs-e-10-0-UI 5', mRNA sequence.
ACCESSION   BM934914
VERSION     BM934914.1  GI:19394066
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 549)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            cDNA Library preparation: Dr. M. Bento Soares, Univeristy of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, Univeristy of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, Univeristy of Iowa
             Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
            Seq primer: M13 REVERSE.
FEATURES             Location/Qualifiers
     source          1. .549
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-BH3-aqs-e-10-0-UI"
                     /clone_lib="NIH_BMAP_M_S4"
                     /dev_stage="27-32 days"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NIH_BMAP_M_S4 library is a subtracted library of a series,
                     ultimately  derived from a mixture of individually tagged
                     normalized libraries from  ten regions of the mouse brain
                     (cerebellum, brain stems, olfactory  bulbs, hypothalamus,
                     cortex, amygdala, basal ganglia, pineal gland,  striatum,
                     hipoccampus) after a series of subtractions to reduce the
                     representation of cDNAs from which ESTs had already been
                     generated. The following serially subtracted libraries
                     were generated in this process: NIH_BMAP_M_S4,
                     NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1,
                     NIH_BMAP_M_S2, NIH_BMAP_M_S1.  The subtracted library
```

(NIH_BMAP_M_S4) was constructed as follows: PCRamplified cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with a pool of the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1 libraries in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (LifeTechnologies) to generate the NIH_BMAP_M_S4 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)"

BASE COUNT 124 a 144 c 149 g 132 t
ORIGIN

Query Match 52.9%; Score 530; DB 14; Length 549;
Best Local Similarity 100.0%; Pred. No.. 1.4e-141;
Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        80 GCAATGCCATTCAGGCCAGGTCTTCCTCCTATAGTGGAGAGTATGGAAGTGGTGGTGGAA 139
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 GCAATGCCATTCAGGCCAGGTCTTCCTCCTATAGTGGAGAGTATGGAAGTGGTGGTGGAA 60

Qy       140 AGCGATTCTCTCATTCTGGCAACCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAG 199
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 AGCGATTCTCTCATTCTGGCAACCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAG 120

Qy       200 TCAACACATACTACATCGTAGGTCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATG 259
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 TCAACACATACTACATCGTAGGTCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATG 180

Qy       260 TGGGTGGTCGCAACGGAGACCTGGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCC 319
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 TGGGTGGTCGCAACGGAGACCTGGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCC 240

Qy       320 AGGTTTCTGGGAAGTACAAGTGGTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCC 379
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 AGGTTTCTGGGAAGTACAAGTGGTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCC 300

Qy       380 GCTATCTGTCTTTTGGGAAAGACAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCA 439
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 GCTATCTGTCTTTTGGGAAAGACAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCA 360

Qy       440 ACACCGTGCTCCGCTTCATCAGTGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGC 499
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 ACACCGTGCTCCGCTTCATCAGTGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGC 420

Qy       500 ACTGGGATGTTTACCCCACTAGCTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGC 559
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 ACTGGGATGTTTACCCCACTAGCTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGC 480

Qy       560 ACTGTGATGAGGAGTAAGAACTCCCTTATCACTAACCCCCATCCAAATGG 609
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 ACTGTGATGAGGAGTAAGAACTCCCTTATCACTAACCCCCATCCAAATGG 530
```

RESULT 6
BI763402
LOCUS BI763402 633 bp mRNA linear EST 25-SEP-2001
DEFINITION 603047590F1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5187790 5', mRNA sequence.
ACCESSION BI763402
VERSION BI763402.1 GI:15754980
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 633)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11469 row: c column: 23
High quality sequence stop: 618.

FEATURES Location/Qualifiers
source 1. .633
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5187790"
/clone_lib="NIH_MGC_116"
/lab_host="DH10B"
/note="Organ: pooled colon, kidney, stomach; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source anonymous pool of 3 colons, age 26 yo male, 49 yo female, 71 yo male colon; 46 yo male kidney, and pool of 2 stomachs, 62 yo male and 70 yo female. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 023. Note: this is a NIH_MGC Library."

BASE COUNT 150 a 169 c 167 g 146 t 1 others
ORIGIN

Query Match 52.1%; Score 521.4; DB 13; Length 633;
Best Local Similarity 99.0%; Pred. No. 4.3e-139;
Matches 566; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

```
Qy        40 ACAGTCGCTCTCCTAGCCCTT-CTCTGTGCCTCAGCCTCTGGCAATGCCATTCAGGCCAG 98
             |||||||||||||||||||| |||||||||||||||||||| | ||||||||||||
Db        13 ACAGTCGCTCTCCTAGCCCTTCCTCTGTGCCTCAGCCTCTGGCAAGG-CATTCAGGCCAG 71

Qy        99 GTCTTCCTCCTATAGTGGAGAGTATGGAAGTGGTGGTGGAAAGCGATTCTCTCATTCTGG 158
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        72 GTCTTCCTCCTATAGTGGAGAGTATGGAAGTGGTGGTGGAAAGCGATTCTCTCATTCTGG 131

Qy       159 CAACCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACACATACTACATCGT 218
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       132 CAACCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACACATACTACATCGT 191

Qy       219 AGGTCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGGTCGCAACGGAGA 278
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       192 AGGTCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGGTCGCAACGGAGA 251

Qy       279 CCTGGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTCTGGGAAGTACAA 338
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       252 CCTGGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTCTGGGAAGTACAA 311

Qy       339 GTGGTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCTGTCTTTTGGGAA 398
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       312 GTGGTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCTGTCTTTTGGGAA 371

Qy       399 AGACAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGTGCTCCGCTTCAT 458
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       372 AGACAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGTGCTCCGCTTCAT 431

Qy       459 CAGTGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGATGTTTACCCCAC 518
             |||||||||||||||||||||||||||||||| |||||||||||||||| ||||||||
Db       432 CAGTGGCCGGTCTGGTTCTCTCATCGATGCCA-TGGCCTGCACTGGGATGTNTACCCCAC 490

Qy       519 TAGCTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGATGAGGAGTAAGA 578
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       491 TAGCTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGATGAGGAGTAAGA 550

Qy       579 ACTCCC-TTATCACTAACCCCCATCCAAATGG 609
```

```
              |||||| ||||||||||||||||||||||||
Db        551 ACTCCCGTTATCACTAACCCCCATCCAAATGG 582


RESULT 7
BM021411
LOCUS       BM021411                572 bp    mRNA    linear   EST 12-MAR-2002
DEFINITION  ie76h03.y1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
            cDNA clone IMAGE:5673053 5' similar to TR:O60844 O60844 HOMOLOG OF
            RAT ZYMOGEN GRANULE MEMBRANE PROTEIN. ;, mRNA sequence.
ACCESSION   BM021411
VERSION     BM021411.1  GI:16535767
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 572)
  AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
            ,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.
            , Jackson,Y. and Bowers,Y.
  TITLE     Endocrine Pancreas Consortium
  JOURNAL   Unpublished (2000)
COMMENT     Other_ESTs: ie76h03.x1
            Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Douglas Melton DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Juliana Brown
            (brown@fas.harvard.edu) This sequence now available from the IMAGE
            consortium, for clone orders contact: info@image.llnl.gov
            High quality sequence stop: 430.
FEATURES             Location/Qualifiers
     source          1. .572
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5673053"
                     /clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
                     /sex="Both"
                     /tissue_type="Islets of Langerhans"
                     /dev_stage="Adult"
                     /lab_host="DH10B",
                     /note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;
                     Site_2: Sal 1; Starting library constructed using
                     SuperScript Plasmid Library kit (Life Technologies). cDNA
                     made by oligo-dT priming. Size-selected by column
                     fractionation; average insert size 1.08 kb. Library was
                     amplified once on solid support and plasmid DNA from
                     library was prepared. The library DNA was normalized by
                     method #4 from Bonaldo, Lennon, and Soares 1996 Genome
                     Research 6:791-806; 0.5 microgram single-stranded library
                     plasmid DNA was mixed with 5 micrograms PCR product
                     representing library inserts and hybridized to an Ecot of
                     20. Single-stranded (unhybridized) plasmids were isolated
                     by hydroxyapatite chromatography and used to make this
                     library."
BASE COUNT      118 a    154 c    155 g    145 t
ORIGIN

  Query Match           51.7%;  Score 517.8;  DB 13;  Length 572;
  Best Local Similarity  99.6%;  Pred. No. 4.5e-138;
  Matches  519;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         27 CCCCAGAATGTTGACAGTCGCTCTCCTAGCCCTTCTCTGTGCCTCAGCCTCTGGCAATGC 86
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         52 CCCCAGAATGTTGACAGTCGCTCTCCTAGCCCTTCTCTGTGCCTCAGCCTCTGGCAATGC 111

Qy         87 CATTCAGGCCAGGTCTTCCTCCTATAGTGGAGAGTATGGAAGTGGTGGTGGAAAGCGATT 146
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        112 CATTCAGGCCAGGTCTTCCTCCTATAGTGGAGAGTATGGAAGTGGTGGTGGAAAGCGATT 171

Qy        147 CTCTCATTCTGGCAACCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACAC 206
              |||||||||||||||||||||||| | ||||||||||||||||||||||||||||||||
Db        172 CTCTCATTCTGGCAACCAGTTGGACGGTCTCATCACCGCCCTCCGGGTCCGAGTCAACAC 231

Qy        207 ATACTACATCGTAGGTCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGG 266
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        232 ATACTACATCGTAGGTCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGG 291

Qy        267 TCGCAACGGAGACCTGGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTC 326
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        292 TCGCAACGGAGACCTGGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTC 351

Qy        327 TGGGAAGTACAAGTGGTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCT 386
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        352 TGGGAAGTACAAGTGGTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCT 411

Qy        387 GTCTTTTGGGAAAGACAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGT 446
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        412 GTCTTTTGGGAAAGACAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGT 471

Qy        447 GCTCCGCTTCATCAGTGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGA 506
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        472 GCTCCGCTTCATCAGTGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGA 531

Qy        507 TGTTTACCCCACTAGCTGCAGCAGATGCTGAGCCTCCTCTC 547
              |||||||||||||||||||||||||||||||||||||||||
Db        532 TGTTTACCCCACTAGCTGCAGCAGATGCTGAGCCTCCTCTC 572


RESULT 8
AI991154/c
LOCUS       AI991154                566 bp    mRNA    linear   EST 27-OCT-1999
DEFINITION  wu38h12.x1 Soares_Dieckgraefe_colon_NHCD Homo sapiens cDNA clone
            IMAGE:2522375 3' similar to TR:O60844 O60844 HOMOLOG OF RAT ZYMOGEN
            GRANULE MEMBRANE PROTEIN. ;, mRNA sequence.
ACCESSION   AI991154
VERSION     AI991154.1  GI:5838059
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 566)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40UP from Gibco
            High quality sequence stop: 328.
FEATURES             Location/Qualifiers
     source          1. .566
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2522375"
                     /clone_lib="Soares_Dieckgraefe_colon_NHCD"
                     /tissue_type="colonic mucosa from 3 patients with Crohn's
                     disease"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCGTCTTTTTTTTTTTTTTTTT 3'],
```

double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Tissue samples provided by Dr. Brian Dieckgraefe (Washington University, dieck@im.wustl.edu); colonic mucosa represents a range of disease involvement from moderate to severe Crohn's disease; samples include both perforating (fistulas) and non-perforating samples. Library constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 140 a 152 c 147 g 127 t
ORIGIN

Query Match 49.7%; Score 497.8; DB 9; Length 566;
Best Local Similarity 96.5%; Pred. No. 2.6e-132;
Matches 520; Conservative 0; Mismatches 17; Indels 2; Gaps 1;

```
Qy        73 GCCTCTGGCAATGCCATTCAGGCCAGGTCTTCCTCCTATAGTGGAGAGTATGGAAGTGGT 132
             ||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||||
Db       566 GCCTCTGGCAATGCATCTCAGGCCAGGTCTTCCTCCTATAGTGGAGAGTATGGAAGTGGT 507

Qy       133 GGTGGAAAGCGATTCTCTCATTCTGGCAACCAGTTGGACGGCCCCATCACCGCCCTCCGG 192
             ||||||||||||||||||||| |||||||||||  |||||||| ||||||||||||||||
Db       506 GGTGGAAAGCGATTCTCTCATTTTGGCAACCAGTGTGACGGCCCTATCACCGCCCTCCGG 447

Qy       193 GTCCGAGTCAACACATACTACATCGTAGGTCTTCAGGTGCGCTATGGCAAGGTGTGGAGC 252
             ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       446 GTCTGAGTCAACACATACTACATCGTAGGTCTTCAGGTGCGCTATGGCAAGGTGTGGAGC 387

Qy       253 GACTATGTGGGTGGTCGCAACGGAGACCTGGAGGAGATCTTTCTGCACCCTGGGGAATCA 312
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       386 GACTATGTGGGTGGTCGCAACGGAGACCTGGAGGAGATCTTTCTGCACCCTGGGGAATCA 327

Qy       313 GTGATCCAGGTTTCTGGGAAGTACAAGTGGTACCTGAAGAAGCTGG--TATTTGTGACAG 370
             ||||||||||||||||||||||||||||||||||||||||||||||  || ||  ||
Db       326 GTGATCCAGGTTTCTGGGAAGTACAAGTGGTACCTGAAGAAGCTGGGTATTTGGTACCAA 267

Qy       371 ACAAGGGCCGCTATCTGTCTTTTGGGAAAGACAGTGGCACAAGTTTCAATGCCGTCCCCT 430
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       266 CCAAGGCCCGCTATCTGTCTTTTGGGAAAGACAGTGGCACAAGTTTCAATGCCGTCCCCT 207

Qy       431 TGCACCCCAACACCGTGCTCCGCTTCATCAGTGGCCGGTCTGGTTCTCTCATCGATGCCA 490
             |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       206 TGCACCCCAACACCGTGCTCCGTTTCATCAGTGGCCGGTCTGGTTCTCTCATCGATGCCA 147

Qy       491 TTGGCCTGCACTGGGATGTTTACCCCACTAGCTGCAGCAGATGCTGAGCCTCCTCTCCTT 550
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       146 TTGGCCTGCACTGGGATGTTTACCCCACTAGCTGCAGCAGATGCTGAGCCTCCTCTCCTT 87

Qy       551 GGCAGGGGCACTGTGATGAGGAGTAAGAACTCCCTTATCACTAACCCCCATCCAAATGG 609
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        86 GGCAGGGGCACTGTGATGAGGAGTAAGAACTCCCTTATCACTAACCCCCATCCAAATGG 28
```

RESULT 9
BQ028047/c
LOCUS BQ028047 520 bp mRNA linear EST 27-MAR-2002
DEFINITION UI-H-CO0-arh-d-06-0-UI.s1 NCI_CGAP_Sub9 Homo sapiens cDNA clone IMAGE:3106570 3', mRNA sequence.
ACCESSION BQ028047
VERSION BQ028047.1 GI:19763326
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 520)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Jose Mercuende
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Seq primer: M13 FORWARD
POLYA=Yes.
FEATURES Location/Qualifiers
source 1. .520
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3106570"
/clone_lib="NCI_CGAP_Sub9"
/tissue_type="mixed"
/dev_stage="mixed"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; tissues: Cholonic mucosa with Crohns disease, Cholonic mucosa with ulcerative colitis, Fetal thymus, Cervix, Cervical adenosquamous carcinoma, Ligament cells, Prostate carcinoma, Bladder carcinoma, Brain oligodenroga ; NCI_CGAP_Sub9 is a subtracted cDNA library constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tags for this library are CGTC, AACG, GGGCC, GGAAG, TAGC, TAAGC, ATGG, AGACA, ATCAC. For additional information, contact: Bento Soares, bento-soares@uiowa.edu
TAG_LIB=UI-H-CO0
TAG_TISSUE=colonic mucosa with Crohns disease
TAG_SEQ=CGTC"
BASE COUNT 119 a 134 c 136 g 131 t
ORIGIN

Query Match 46.9%; Score 469.4; DB 14; Length 520;
Best Local Similarity 99.8%; Pred. No. 3.8e-124;
Matches 470; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       139 AAGCGATTCTCTCATTCTGGCAACCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGA 198
             | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       515 AGGCGATTCTCTCATTCTGGCAACCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGA 456

Qy       199 GTCAACACATACTACATCGTAGGTCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTAT 258
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       455 GTCAACACATACTACATCGTAGGTCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTAT 396

Qy       259 GTGGGTGGTCGCAACGGAGACCTGGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATC 318
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       395 GTGGGTGGTCGCAACGGAGACCTGGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATC 336

Qy       319 CAGGTTTCTGGGAAGTACAAGTGGTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGC 378
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       335 CAGGTTTCTGGGAAGTACAAGTGGTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGC 276

Qy       379 CGCTATCTGTCTTTTGGGAAAGACAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCC 438
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       275 CGCTATCTGTCTTTTGGGAAAGACAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCC 216

Qy       439 AACACCGTGCTCCGCTTCATCAGTGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTG 498
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       215 AACACCGTGCTCCGCTTCATCAGTGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTG 156

Qy       499 CACTGGGATGTTTACCCCACTAGCTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGG 558
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       155 CACTGGGATGTTTACCCCACTAGCTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGG 96
```

```
Qy        559 CACTGTGATGAGGAGTAAGAACTCCCTTATCACTAACCCCCATCCAAATGG 609
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         95 CACTGTGATGAGGAGTAAGAACTCCCTTATCACTAACCCCCATCCAAATGG 45


RESULT 10
BM987195/c
LOCUS       BM987195                490 bp    mRNA    linear   EST 25-MAR-2002
DEFINITION  UI-H-CO0-aql-a-10-0-UI.s1 NCI_CGAP_Sub9 Homo sapiens cDNA clone
            IMAGE: 3104154 3', mRNA sequence.
ACCESSION   BM987195
VERSION     BM987195.1  GI:19706584
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 490)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jose Mercuende
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .490
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE: 3104154"
                     /clone_lib="NCI_CGAP_Sub9"
                     /tissue_type="mixed"
                     /dev_stage="mixed"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3-Pac (Pharmacia) with a modified
                     polylinker; Site_1: EcoR I; Site_2: Not I; tissues:
                     Cholonic mucosa with Crohns disease, Cholonic mucosa with
                     ulcerative colitis, Fetal thymus, Cervix, Cervical
                     adenosquamous carcinoma, Ligament cells, Prostate
                     carcinoma, Bladder carcinoma, Brain oligodenroga ;
                     NCI_CGAP_Sub9 is a subtracted cDNA library constructed
                     according to Bonaldo, Lennon and Soares, Genome Research,
                     6:791-806, 1996. First strand cDNA synthesis was primed
                     with an oligo-dT primer containing a Not I site. Double
                     stranded cDNA was ligated to an EcoR I adaptor, digested
                     with Not I, and cloned directionally into pT7T3-Pac
                     vector. The oligonucleotide used to prime the synthesis of
                     first-strand cDNA contains a library tag sequence that is
                     located between the Not I site and the (dT)18 tail. The
                     sequence tags for this library are CGTC, AACG, GGGCC,
                     GGAAG, TAGC, TAAGC, ATGG, AGACA, ATCAC. For additional
                     information, contact: Bento Soares, bento-soares@uiowa.edu
                     TAG_LIB=UI-H-CO0
                     TAG_TISSUE=colonic mucosa with Crohns disease
                     TAG_SEQ=CGTC"
BASE COUNT      111 a    127 c    127 g    125 t
ORIGIN

  Query Match           44.7%;  Score 447;  DB 14;  Length 490;
  Best Local Similarity  100.0%;  Pred. No. 1e-117;
  Matches  447;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        163 CAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACACATACTACATCGTAGGT 222
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        490 CAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACACATACTACATCGTAGGT 431

Qy        223 CTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGGTCGCAACGGAGACCTG 282
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        430 CTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGGTCGCAACGGAGACCTG 371

Qy        283 GAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTCTGGGAAGTACAAGTGG 342
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        370 GAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTCTGGGAAGTACAAGTGG 311

Qy        343 TACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCTGTCTTTTGGGAAAGAC 402
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        310 TACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCTGTCTTTTGGGAAAGAC 251

Qy        403 AGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGTGCTCCGCTTCATCAGT 462
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        250 AGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGTGCTCCGCTTCATCAGT 191

Qy        463 GGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGATGTTTACCCCACTAGC 522
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        190 GGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGATGTTTACCCCACTAGC 131

Qy        523 TGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGATGAGGAGTAAGAACTC 582
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        130 TGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGATGAGGAGTAAGAACTC 71

Qy        583 CCTTATCACTAACCCCCATCCAAATGG 609
              ||||||||||||||||||||||||||
Db         70 CCTTATCACTAACCCCCATCCAAATGG 44


RESULT 11
BM987408/c
LOCUS       BM987408                485 bp    mRNA    linear   EST 17-JUN-2002
DEFINITION  UI-H-CO0-asg-a-06-0-UI.s1 NCI_CGAP_Sub9 Homo sapiens cDNA clone
            IMAGE:5858578 3', mRNA sequence.
ACCESSION   BM987408
VERSION     BM987408.1  GI:19706797
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 485)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jose Mercuende
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .485
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5858578"
                     /clone_lib="NCI_CGAP_Sub9"
                     /tissue_type="mixed"
                     /dev_stage="mixed"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3-Pac (Pharmacia) with a modified
                     polylinker; Site_1: EcoR I; Site_2: Not I; tissues:
                     Cholonic mucosa with Crohns disease, Cholonic mucosa with
                     ulcerative colitis, Fetal thymus, Cervix, Cervical
                     adenosquamous carcinoma, Ligament cells, Prostate
                     carcinoma, Bladder carcinoma, Brain oligodenroga ;
                     NCI_CGAP_Sub9 is a subtracted cDNA library constructed
                     according to Bonaldo, Lennon and Soares, Genome Research,
```

6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tags for this library are CGTC, AACG, GGGCC, GGAAG, TAGC, TAAGC, ATGG, AGACA, ATCAC. For additional information, contact: Bento Soares, bento-soares@uiowa.edu
TAG_LIB=UI-H-CO0
TAG_TISSUE=colonic mucosa with Crohns disease
TAG_SEQ=CGTC"
BASE COUNT 110 a 125 c 126 g 124 t
ORIGIN

Query Match 44.1%; Score 441; DB 14; Length 485;
Best Local Similarity 100.0%; Pred. No. 5.5e-116;
Matches 441; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      169 GACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACACATACTACATCGTAGGTCTTCAG 228
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      485 GACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACACATACTACATCGTAGGTCTTCAG 426

Qy      229 GTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGGTCGCAACGGAGACCTGGAGGAG 288
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      425 GTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGGTCGCAACGGAGACCTGGAGGAG 366

Qy      289 ATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTCTGGGAAGTACAAGTGGTACCTG 348
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      365 ATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTCTGGGAAGTACAAGTGGTACCTG 306

Qy      349 AAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCTGTCTTTTGGGAAAGACAGTGGC 408
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      305 AAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCTGTCTTTTGGGAAAGACAGTGGC 246

Qy      409 ACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGTGCTCCGCTTCATCAGTGGCCGG 468
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      245 ACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGTGCTCCGCTTCATCAGTGGCCGG 186

Qy      469 TCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGATGTTTACCCCACTAGCTGCAGC 528
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      185 TCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGATGTTTACCCCACTAGCTGCAGC 126

Qy      529 AGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGATGAGGAGTAAGAACTCCCTTAT 588
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      125 AGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGATGAGGAGTAAGAACTCCCTTAT 66

Qy      589 CACTAACCCCCATCCAAATGG 609
            |||||||||||||||||||||
Db       65 CACTAACCCCCATCCAAATGG 45
```

RESULT 12
BM987303/c
LOCUS BM987303 492 bp mRNA linear EST 25-MAR-2002
DEFINITION UI-H-CO0-aqn-e-10-0-UI.s1 NCI_CGAP_Sub9 Homo sapiens cDNA clone IMAGE: 3104706 3', mRNA sequence.
ACCESSION BM987303
VERSION BM987303.1 GI:19706692
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 492)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Jose Mercuende
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Seq primer: M13 FORWARD
POLYA=Yes.
FEATURES Location/Qualifiers
source 1. .492
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE: 3104706"
/clone_lib="NCI_CGAP_Sub9"
/tissue_type="mixed"
/dev_stage="mixed"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; tissues: Cholonic mucosa with Crohns disease, Cholonic mucosa with ulcerative colitis, Fetal thymus, Cervix, Cervical adenosquamous carcinoma, Ligament cells, Prostate carcinoma, Bladder carcinoma, Brain oligodenroga ; NCI_CGAP_Sub9 is a subtracted cDNA library constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tags for this library are CGTC, AACG, GGGCC, GGAAG, TAGC, TAAGC, ATGG, AGACA, ATCAC. For additional information, contact: Bento Soares, bento-soares@uiowa.edu
TAG_LIB=UI-H-CO0
TAG_TISSUE=colonic mucosa with Crohns disease
TAG_SEQ=CGTC"
BASE COUNT 112 a 127 c 127 g 126 t
ORIGIN

Query Match 43.5%; Score 435.4; DB 14; Length 492;
Best Local Similarity 99.8%; Pred. No. 2.3e-114;
Matches 436; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      173 GCCCCATCACCGCCCTCCGGGTCCGAGTCAACACATACTACATCGTAGGTCTTCAGGTGC 232
            | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481 GGCCCATCACCGCCCTCCGGGTCCGAGTCAACACATACTACATCGTAGGTCTTCAGGTGC 422

Qy      233 GCTATGGCAAGGTGTGGAGCGACTATGTGGGTGGTCGCAACGGAGACCTGGAGGAGATCT 292
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421 GCTATGGCAAGGTGTGGAGCGACTATGTGGGTGGTCGCAACGGAGACCTGGAGGAGATCT 362

Qy      293 TTCTGCACCCTGGGGAATCAGTGATCCAGGTTTCTGGGAAGTACAAGTGGTACCTGAAGA 352
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 TTCTGCACCCTGGGGAATCAGTGATCCAGGTTTCTGGGAAGTACAAGTGGTACCTGAAGA 302

Qy      353 AGCTGGTATTTGTGACAGACAAGGGCCGCTATCTGTCTTTTGGGAAAGACAGTGGCACAA 412
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 AGCTGGTATTTGTGACAGACAAGGGCCGCTATCTGTCTTTTGGGAAAGACAGTGGCACAA 242

Qy      413 GTTTCAATGCCGTCCCCTTGCACCCCAACACCGTGCTCCGCTTCATCAGTGGCCGGTCTG 472
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 GTTTCAATGCCGTCCCCTTGCACCCCAACACCGTGCTCCGCTTCATCAGTGGCCGGTCTG 182

Qy      473 GTTCTCTCATCGATGCCATTGGCCTGCACTGGGATGTTTACCCCACTAGCTGCAGCAGAT 532
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 GTTCTCTCATCGATGCCATTGGCCTGCACTGGGATGTTTACCCCACTAGCTGCAGCAGAT 122

Qy      533 GCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGATGAGGAGTAAGAACTCCCTTATCACT 592
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 GCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGATGAGGAGTAAGAACTCCCTTATCACT 62

Qy      593 AACCCCCATCCAAATGG 609
```

```
                  ||||||||||||||||
Db          61 AACCCCCATCCAAATGG 45


RESULT 13
AA828206/c
LOCUS       AA828206                497 bp    mRNA    linear   EST 20-FEB-1998
DEFINITION  of05d11.s1 NCI_CGAP_Co12 Homo sapiens cDNA clone IMAGE:1420245 3'
            similar to TR:Q63680 Q63680 ZG-16P. ;, mRNA sequence.
ACCESSION   AA828206
VERSION     AA828206.1  GI:2900569
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 497)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: Stratagene, Inc.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 339.
FEATURES             Location/Qualifiers
     source          1. .497
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1420245"
                     /clone_lib="NCI_CGAP_Co12"
                     /sex="mixed"
                     /tissue_type="colon tumor"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: colon; Vector: Bluescript SK-; Site_1: EcoRI
                     ; Site_2: XhoI; Cloned unidirectionally.  Primer: Oligo
                     dT. Pooled colon tumors. 5' adaptor sequence: 5'
                     GAATTCGGCACGAG 3' 3' adaptor sequence: 5'
                     CTCGAGTTTTTTTTTTTTTTTT 3' Average insert size:  1.2 kb."
BASE COUNT      120 a    131 c    131 g    115 t
ORIGIN

  Query Match           43.5%;  Score 435.2;  DB 9;  Length 497;
  Best Local Similarity   97.7%;  Pred. No. 2.6e-114;
  Matches  463;  Conservative    0;  Mismatches    8;  Indels    3;  Gaps    2;

Qy        136 GGAAAGCGATTCTCTCATTCTGGCAACCAGTTGGACGGCCCCATCACCGCCCTCCGGGTC 195
              ||||| |||||||||||||||||||||| |||| |||  |||||| |   ||||||
Db        497 GGAAAACGATTCTCTCATTCTGGCAACCAGTT-GACGCCCCATTCACCGGC--TCGGGTC 441

Qy        196 CGAGTCAACACATACTACATCGTAGGTCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGAC 255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        440 CGAGTCAACACATACTACATCGTAGGTCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGAC 381

Qy        256 TATGTGGGTGGTCGCAACGGAGACCTGGAGGAGATCTTTCTGCACCCTGGGGAATCAGTG 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        380 TATGTGGGTGGTCGCAACGGAGACCTGGAGGAGATCTTTCTGCACCCTGGGGAATCAGTG 321

Qy        316 ATCCAGGTTTCTGGGAAGTACAAGTGGTACCTGAAGAAGCTGGTATTTGTGACAGACAAG 375
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        320 ATCCAGGTTTCTGGGAAGTACAAGTGGTACCTGAAGAAGCTGGTATTTGTGACAGACAAG 261

Qy        376 GGCCGCTATCTGTCTTTTGGGAAAGACAGTGGCACAAGTTTCAATGCCGTCCCCTTGCAC 435
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        260 GGCCGCTATCTGTCTTTTGGGAAAGACAGTGGCACAAGTTTCAATGCCGTCCCCTTGCAC 201

Qy        436 CCCAACACCGTGCTCCGCTTCATCAGTGGCCGGTCTGGTTCTCTCATCGATGCCATTGGC 495
              |||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||
Db        200 CCCAACACCGTCGTCCGCTTCATCAGTGGCCGGTCTGGTTCTCTCATCGATGCCATTGGC 141

Qy        496 CTGCACTGGGATGTTTACCCCACTAGCTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAG 555
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        140 CTGCACTGGGATGTTTACCCCACTAGCTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAG 81

Qy        556 GGGCACTGTGATGAGGAGTAAGAACTCCCTTATCACTAACCCCCATCCAAATGG 609
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         80 GGGCACTGTGATGAGGAGTAAGAACTCCCTTATCACTAACCCCCATCCAAATGG 27


RESULT 14
AI749030/c
LOCUS       AI749030                455 bp    mRNA    linear   EST 22-JUN-1999
DEFINITION  at38b05.x1 Barstead colon HPLRB7 Homo sapiens cDNA clone
            IMAGE:2374257 3' similar to TR:O60844 O60844 HOMOLOG OF RAT ZYMOGEN
            GRANULE MEMBRANE PROTEIN. ;, mRNA sequence.
ACCESSION   AI749030
VERSION     AI749030.1  GI:5127229
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 455)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
            ,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Possible reversed clone: similarity on wrong strand
            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .455
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2374257"
                     /clone_lib="Barstead colon HPLRB7"
                     /sex="male"
                     /dev_stage="adult, age 25"
                     /lab_host="DH10B (phage resistant)"
                     /note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     [5' AATTCACTAGTAAT 3' and 5' ATTACTAGTG 3'], digested
                     with Not I and cloned into the Not I and Eco RI sites of
                     the modified pT7T3 vector. Library constructed by Bob
                     Barstead."
BASE COUNT      109 a    123 c    120 g    102 t      1 others
ORIGIN

  Query Match           42.6%;  Score 426.4;  DB 9;  Length 455;
  Best Local Similarity   99.5%;  Pred. No. 8.6e-112;
  Matches  427;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        181 ACCGCCCTCCGGGTCCGAGTCAACACATACTACATCGTAGGTCTTCAGGTGCGCTATGGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        455 ACCGCCCTCCGGGTCCGAGTCAACACATACTACATCGTAGGTCTTCAGGTGCGCTATGGC 396
```

```
Qy        241 AAGGTGTGGAGCGACTATGTGGGTGGTCGCAACGGAGACCTGGAGGAGATCTTTCTGCAC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        395 AAGGTGTGGAGCGACTATGTGGGTGGTCGCAACGGAGACCTGGAGGAGATCTTTCTGCAC 336

Qy        301 CCTGGGGAATCAGTGATCCAGGTTTCTGGGAAGTACAAGTGGTACCTGAAGAAGCTGGTA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        335 CCTGGGGAATCAGTGATCCAGGTTTCTGGGAAGTACAAGTGGTACCTGAAGAAGCTGGTA 276

Qy        361 TTTGTGACAGACAAGGGCCGCTATCTGTCTTTTGGGAAAGACAGTGGCACAAGTTTCAAT 420
              |||||||||| || ||||||||||||||||||||||||||||||||||||||||||||||
Db        275 TTTGTGACAGNCAGGGGCCGCTATCTGTCTTTTGGGAAAGACAGTGGCACAAGTTTCAAT 216

Qy        421 GCCGTCCCCTTGCACCCCAACACCGTGCTCCGCTTCATCAGTGGCCGGTCTGGTTCTCTC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        215 GCCGTCCCCTTGCACCCCAACACCGTGCTCCGCTTCATCAGTGGCCGGTCTGGTTCTCTC 156

Qy        481 ATCGATGCCATTGGCCTGCACTGGGATGTTTACCCCACTAGCTGCAGCAGATGCTGAGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        155 ATCGATGCCATTGGCCTGCACTGGGATGTTTACCCCACTAGCTGCAGCAGATGCTGAGCC 96

Qy        541 TCCTCTCCTTGGCAGGGGCACTGTGATGAGGAGTAAGAACTCCCTTATCACTAACCCCCA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         95 TCCTCTCCTTGGCAGGGGCACTGTGATGAGGAGTAAGAACTCCCTTATCACTAACCCCCA 36

Qy        601 TCCAAATGG 609
              |||||||||
Db         35 TCCAAATGG 27
```

RESULT 15
AW001526/c
LOCUS AW001526 467 bp mRNA linear EST 09-MAR-2000
DEFINITION wu34a06.x1 Soares_Dieckgraefe_colon_NHCD Homo sapiens cDNA clone IMAGE:2521906 3' similar to TR:O60844 O60844 HOMOLOG OF RAT ZYMOGEN GRANULE MEMBRANE PROTEIN. ;, mRNA sequence.
ACCESSION AW001526
VERSION AW001526.1 GI:5848442
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 467)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 610 Std Error: 0.00
Seq primer: -40UP from Gibco.
FEATURES Location/Qualifiers
source 1. .467
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2521906"
/clone_lib="Soares_Dieckgraefe_colon_NHCD"
/tissue_type="colonic mucosa from 3 patients with Crohn's disease"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCCGTCTTTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Tissue samples provided by Dr. Brian Dieckgraefe (Washington University, dieck@im.wustl.edu); colonic mucosa represents a range of disease involvement from moderate to severe Crohn's disease; samples include both perforating (fistulas) and non-perforating samples. Library constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 112 a 125 c 125 g 105 t
ORIGIN

Query Match 42.3%; Score 423.2; DB 10; Length 467;
Best Local Similarity 99.1%; Pred. No. 7.3e-111;
Matches 436; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

```
Qy        171 CGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACACATACTACATCGTAGGTCTTCAGGT 230
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        467 CGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACACATATTACATCGTAGGTCTTCAGGT 408

Qy        231 GCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGGTCGCAACGGAGACCTGGAGGAGAT 290
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        407 GCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGGTCGCAACGGAGACCTGGAGGAGAT 348

Qy        291 CTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTCTGGGAAGTACAAGTGGTACCTGAA 350
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        347 CTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTCTGGGAAGTACAAGTGGTACTTGAA 288

Qy        351 GAAGCTGGTATTTGTGACAGACAA-GGGCCGCTATCTGTCTTTTGGGAAAGACAGTGGCA 409
              |||||||||||||||||||| ||| ||||||||||||||||||||||||||||||||||||
Db        287 GAAGCTGGTATTTGTGACAGCCAAGGGGCCGCTATCTGTCTTTTGGGAAAGACAGTGGCA 228

Qy        410 CAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGTGCTCCGCTTCATCAGTGGCCGGT 469
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        227 CAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGTGCTCCGCTTCATCAGTGGCCGGT 168

Qy        470 CTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGATGTTTACCCCACTAGCTGCAGCA 529
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        167 CTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGATGTTTACCCCACTAGCTGCAGCA 108

Qy        530 GATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGATGAGGAGTAAGAACTCCCTTATC 589
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        107 GATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGATGAGGAGTAAGAACTCCCTTATC 48

Qy        590 ACTAACCCCCATCCAAATGG 609
              ||||||||||||||||||||
Db         47 ACTAACCCCCATCCAAATGG 28
```

Search completed: July 16, 2003, 14:55:35
Job time : 1097.78 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 16, 2003, 10:35:27 ; Search time 39.0516 Seconds
(without alignments)
7860.974 Million cell updates/sec

Title: US-09-981-353-24
Perfect score: 1001
Sequence: 1 atgattacgaattcgagctc.....gtgagcaaaaggccagcaaa 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	545.8	54.5	560	1	US-08-469-667-10	Sequence 10, Appl
	2	545.8	54.5	560	4	US-09-224-110-10	Sequence 10, Appl
	3	545.8	54.5	560	5	PCT-US95-07289-10	Sequence 10, Appl
	4	352.2	35.2	3789	4	US-09-075-019-8	Sequence 8, Appli
	5	352	35.2	3983	4	US-09-481-049-1	Sequence 1, Appli
	6	351.8	35.1	4824	2	US-08-485-139-5	Sequence 5, Appli
	7	351.8	35.1	4824	3	US-08-750-357-5	Sequence 5, Appli
c	8	351.8	35.1	8710	1	US-08-480-882B-3	Sequence 3, Appli
c	9	351.8	35.1	8710	1	US-08-480-210-3	Sequence 3, Appli
c	10	351.8	35.1	9019	1	US-08-480-882B-4	Sequence 4, Appli
c	11	351.8	35.1	9019	1	US-08-480-210-4	Sequence 4, Appli
	12	350.8	35.0	738	1	US-08-726-462B-3	Sequence 3, Appli
	13	350.8	35.0	738	2	US-09-046-203-3	Sequence 3, Appli
	14	350.8	35.0	738	3	US-09-272-104-3	Sequence 3, Appli
	15	350.8	35.0	738	4	US-09-272-097-3	Sequence 3, Appli
	16	350.8	35.0	4283	1	US-08-343-401A-3	Sequence 3, Appli
	17	350.8	35.0	4283	1	US-08-445-265A-1	Sequence 1, Appli
	18	350.8	35.0	4283	3	US-08-990-442-1	Sequence 1, Appli
c	19	350.8	35.0	10306	3	US-08-716-351A-4	Sequence 4, Appli
c	20	350.8	35.0	10970	3	US-08-716-351A-5	Sequence 5, Appli
c	21	350.6	35.0	5692	4	US-09-526-993-11	Sequence 11, Appl
c	22	350.6	35.0	5737	4	US-09-526-993-9	Sequence 9, Appli
c	23	350.6	35.0	6157	4	US-09-526-993-10	Sequence 10, Appl
c	24	350.6	35.0	6202	4	US-09-526-993-8	Sequence 8, Appli
	25	350.2	35.0	716	4	US-08-998-416-55	Sequence 55, Appl
	26	350.2	35.0	752	4	US-08-976-259-108	Sequence 108, App
	27	350.2	35.0	2961	4	US-08-446-935-6	Sequence 6, Appli
	28	350.2	35.0	3699	3	US-08-646-538-6	Sequence 6, Appli
	29	350.2	35.0	3699	4	US-09-503-222-6	Sequence 6, Appli
c	30	350.2	35.0	4145	3	US-08-651-472-62	Sequence 62, Appl
c	31	350.2	35.0	4145	4	US-08-358-928-62	Sequence 62, Appl
c	32	350.2	35.0	4277	3	US-08-651-472-63	Sequence 63, Appl
c	33	350.2	35.0	4277	4	US-08-358-928-63	Sequence 63, Appl
c	34	350.2	35.0	4435	2	US-08-792-824-1	Sequence 1, Appli
c	35	350.2	35.0	4525	2	US-08-613-861-2	Sequence 2, Appli
	36	350.2	35.0	4539	1	US-08-119-512-1	Sequence 1, Appli
	37	350.2	35.0	4539	1	US-08-488-015B-1	Sequence 1, Appli
	38	350.2	35.0	4542	3	US-08-814-412-11	Sequence 11, Appl
c	39	350.2	35.0	4883	4	US-09-608-730B-21	Sequence 21, Appl
	40	350.2	35.0	4965	3	US-08-675-566-22	Sequence 22, Appl
	41	350.2	35.0	5033	1	US-08-038-768A-1	Sequence 1, Appli
	42	350.2	35.0	5109	3	US-08-675-566-19	Sequence 19, Appl
	43	350.2	35.0	5178	2	US-08-474-169-2	Sequence 2, Appli
	44	350.2	35.0	5356	4	US-08-446-935-1	Sequence 1, Appli
	45	350.2	35.0	5382	4	US-09-479-122-21	Sequence 21, Appl

ALIGNMENTS

```
RESULT 1
US-08-469-667-10
; Sequence 10, Application US/08469667
; Patent No. 5733748
;  GENERAL INFORMATION:
;    APPLICANT:  Yu, Guo-Liang
;    APPLICANT:  Rosen, Craig
;    TITLE OF INVENTION:  Colon Specific Genes and Proteins
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Carella, Byrne, Bain, Gilfillan, Cecchi,
;      ADDRESSEE:  Stewart & Olstein
;      STREET:  6 Becker Farm Road
;      CITY:  Roseland
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  07068-1739
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/469,667
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ferraro, Gregory D.
;      REGISTRATION NUMBER:  36,134
;      REFERENCE/DOCKET NUMBER:  325800-435
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-994-1700
;      TELEFAX:  201-994-1744
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  560 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2..490
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  2..490
US-08-469-667-10

  Query Match            54.5%;  Score 545.8;  DB 1;  Length 560;
  Best Local Similarity  99.5%;  Pred. No. 2.7e-158;
```

```
  Matches  558;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy         42 AGTCGCTCTCCTAGCCCTTCTCTGTGCCTCAGCCTCTGGCAATGCCATTCAGGCCAGGTC 101
              ||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db          1 AGTCGCTCTCCTAGCCCTTCTCTGTGCCTCACCCTCTGGCAATGCCATTCAGGCCAGGTC 60

Qy        102 TTCCTCCTATAGTGGAGAGTATGGAAGTGGTGGTGGAAAGCGATTCTCTCATTCTGGCAA 161
              ||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db         61 TTCCTCCTATAGTGGAGAGTATGGAGGTGGTGGTGGAAAGCGATTCTCTCATTCTGGCAA 120

Qy        162 CCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACACATACTACATCGTAGG 221
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACACATACTACATCGTAGG 180

Qy        222 TCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGGTCGCAACGGAGACCT 281
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGGTCGCAACGGAGACCT 240

Qy        282 GGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTCTGGGAAGTACAAGTG 341
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTCTGGGAAGTACAAGTG 300

Qy        342 GTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCTGTCTTTTGGGAAAGA 401
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCTGTCTTTTGGGAAAGA 360

Qy        402 CAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGTGCTCCGCTTCATCAG 461
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGTGCTCCGCTTCATCAG 420

Qy        462 TGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGATGTTTACCCCACTAG 521
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGATGTTTACCCCACTAG 480

Qy        522 CTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGATGAGGAGTAAGAACT 581
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGATGAGGAGTAAGAACT 540

Qy        582 CCCTTATCACTAACCCCCATC 602
                |||||||||||||||||||
Db        541 -CCTTATCACTAACCCCCATC 560


RESULT 2
US-09-224-110-10
; Sequence 10, Application US/09224110
; Patent No. 6337195
;  GENERAL INFORMATION:
;    APPLICANT:  Yu, Guo-Liang
;    APPLICANT:  Rosen, Craig
;    TITLE OF INVENTION:  Colon Specific Genes and Proteins
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Carella, Byrne, Bain, Gilfillan, Cecchi,
;      ADDRESSEE:  Stewart & Olstein
;      STREET:  6 Becker Farm Road
;      CITY:  Roseland
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  07068-1739
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/224,110
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/469,667
;      FILING DATE:  06-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ferraro, Gregory D.
;      REGISTRATION NUMBER:  36,134
;      REFERENCE/DOCKET NUMBER:  325800-435
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-994-1700
;      TELEFAX:  201-994-1744
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  560 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2..490
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  2..490
US-09-224-110-10

  Query Match             54.5%;  Score 545.8;  DB 4;  Length 560;
  Best Local Similarity   99.5%;  Pred. No. 2.7e-158;
  Matches  558;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy         42 AGTCGCTCTCCTAGCCCTTCTCTGTGCCTCAGCCTCTGGCAATGCCATTCAGGCCAGGTC 101
              ||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db          1 AGTCGCTCTCCTAGCCCTTCTCTGTGCCTCACCCTCTGGCAATGCCATTCAGGCCAGGTC 60

Qy        102 TTCCTCCTATAGTGGAGAGTATGGAAGTGGTGGTGGAAAGCGATTCTCTCATTCTGGCAA 161
              ||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db         61 TTCCTCCTATAGTGGAGAGTATGGAGGTGGTGGTGGAAAGCGATTCTCTCATTCTGGCAA 120

Qy        162 CCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACACATACTACATCGTAGG 221
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACACATACTACATCGTAGG 180

Qy        222 TCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGGTCGCAACGGAGACCT 281
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGGTCGCAACGGAGACCT 240

Qy        282 GGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTCTGGGAAGTACAAGTG 341
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTCTGGGAAGTACAAGTG 300

Qy        342 GTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCTGTCTTTTGGGAAAGA 401
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCTGTCTTTTGGGAAAGA 360

Qy        402 CAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGTGCTCCGCTTCATCAG 461
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGTGCTCCGCTTCATCAG 420

Qy        462 TGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGATGTTTACCCCACTAG 521
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGATGTTTACCCCACTAG 480

Qy        522 CTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGATGAGGAGTAAGAACT 581
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGATGAGGAGTAAGAACT 540

Qy        582 CCCTTATCACTAACCCCCATC 602
                |||||||||||||||||||
Db        541 -CCTTATCACTAACCCCCATC 560


RESULT 3
PCT-US95-07289-10
; Sequence 10, Application PC/TUS9507289
;  GENERAL INFORMATION:
;    APPLICANT:  Yu, Guo-Liang
;    APPLICANT:  Rosen, Craig
;    TITLE OF INVENTION:  Colon Specific Genes and Proteins
```

```
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Carella, Byrne, Bain, Gilfillan, Cecchi,
;      ADDRESSEE:  Stewart & Olstein
;      STREET:  6 Becker Farm Road
;      CITY:  Roseland
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  07068-1739
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/07289
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ferraro, Gregory D.
;      REGISTRATION NUMBER:  36,134
;      REFERENCE/DOCKET NUMBER:  325800-265
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-994-1700
;      TELEFAX:  201-994-1744
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  560 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2..490
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  2..490
PCT-US95-07289-10

  Query Match           54.5%;  Score 545.8;  DB 5;  Length 560;
  Best Local Similarity  99.5%;  Pred. No. 2.7e-158;
  Matches  558;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy         42 AGTCGCTCTCCTAGCCCTTCTCTGTGCCTCAGCCTCTGGCAATGCCATTCAGGCCAGGTC 101
              ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db          1 AGTCGCTCTCCTAGCCCTTCTCTGTGCCTCACCCTCTGGCAATGCCATTCAGGCCAGGTC 60

Qy        102 TTCCTCCTATAGTGGAGAGTATGGAAGTGGTGGTGGAAAGCGATTCTCTCATTCTGGCAA 161
              ||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db         61 TTCCTCCTATAGTGGAGAGTATGGAGGTGGTGGTGGAAAGCGATTCTCTCATTCTGGCAA 120

Qy        162 CCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACACATACTACATCGTAGG 221
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACACATACTACATCGTAGG 180

Qy        222 TCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGGTCGCAACGGAGACCT 281
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGGTCGCAACGGAGACCT 240

Qy        282 GGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTCTGGGAAGTACAAGTG 341
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTCTGGGAAGTACAAGTG 300

Qy        342 GTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCTGTCTTTTGGGAAAGA 401
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCTGTCTTTTGGGAAAGA 360

Qy        402 CAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGTGCTCCGCTTCATCAG 461
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGTGCTCCGCTTCATCAG 420

Qy        462 TGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGATGTTTACCCCACTAG 521
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGATGTTTACCCCACTAG 480

Qy        522 CTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGATGAGGAGTAAGAACT 581
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGATGAGGAGTAAGAACT 540

Qy        582 CCCTTATCACTAACCCCCATC 602
              ||||||||||||||||||||
Db        541 -CCTTATCACTAACCCCCATC 560


RESULT 4
US-09-075-019-8
; Sequence 8, Application US/09075019
; Patent No. 6190658
;  GENERAL INFORMATION:
;    APPLICANT:  UTC IR459
;    TITLE OF INVENTION:  A GENETICALLY MODIFIED MANGANESE
;    TITLE OF INVENTION:  SUPEROXIDE DISMUTASE FOR TREATING OXIDATIVE DAMAGE
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sheridan Ross P.C.
;      STREET:  1700 Lincoln St., Suite 3500
;      CITY:  Denver
;      STATE:  CO
;      COUNTRY:  USA
;      ZIP:  80203
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/075,019
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kovarik, Joseph E.
;      REGISTRATION NUMBER:  33,005
;      REFERENCE/DOCKET NUMBER:  2848-22
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 863-9700
;      TELEFAX:  (303) 863-0223
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3789 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-075-019-8

  Query Match           35.2%;  Score 352.2;  DB 4;  Length 3789;
  Best Local Similarity  88.4%;  Pred. No. 2.4e-98;
  Matches  357;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        598 CCATCCAAATGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCCTGTGTG 657
              || || |  ||                                          |||||||||
Db       1529 CCGTCGACCTGCAGGGCATGCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTG 1588

Qy        658 AAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGC 717
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1589 AAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGC 1648

Qy        718 CTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTT 777
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1649 CTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTT 1708

Qy        778 CCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGG 837
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1709 CCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGG 1768
```

```
Qy        838 CGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGT 897
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1769 CGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGT 1828

Qy        898 TCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATC 957
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1829 TCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATC 1888

Qy        958 AGGGGGATACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1001
              |||||   ||||||||||||||||||||||||||||||||||||
Db       1889 AGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1932


RESULT 5
US-09-481-049-1
; Sequence 1, Application US/09481049
; Patent No. 6395485
; GENERAL INFORMATION:
;  APPLICANT: DE BEUCKELEER, Marc
;  TITLE OF INVENTION: METHODS FOR IDENTIFYING ELITE EVENT GAT-ZM1 IN
;  TITLE OF INVENTION: BIOLOGICAL SAMPLES
;  FILE REFERENCE: 514412-2025
;  CURRENT APPLICATION NUMBER: US/09/481,049
;  CURRENT FILING DATE: 2000-01-11
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 3983
;   TYPE: DNA
;   ORGANISM: Zea mays
US-09-481-049-1

  Query Match           35.2%;  Score 352;  DB 4;  Length 3983;
  Best Local Similarity  88.0%;  Pred. No. 2.8e-98;
  Matches  358;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy        595 CCCCCATCCAAATGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCCTGT 654
              || | ||   |||                                          ||||||
Db       1720 CCTCTATTTGAATCTTTGACTCCATGGGAATTCGTAATCATGGTCATAGCTGTTTCCTGT 1779

Qy        655 GTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAA 714
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1780 GTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAA 1839

Qy        715 AGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGC 774
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1840 AGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGC 1899

Qy        775 TTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAG 834
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1900 TTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAG 1959

Qy        835 AGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGT 894
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1960 AGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGT 2019

Qy        895 CGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGA 954
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2020 CGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGA 2079

Qy        955 ATCAGGGGGATACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1001
              ||||||||  ||||||||||||||||||||||||||||||||||||
Db       2080 ATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 2126


RESULT 6
US-08-485-139-5
; Sequence 5, Application US/08485139
; Patent No. 5880331
;  GENERAL INFORMATION:
;    APPLICANT:  KREBBERS, Enno
;    APPLICANT:  WILLIAMS, Mark
;    APPLICANT:  LEEMANS, Jan
;    TITLE OF INVENTION:  USE OF ANTHOCYANIN GENES TO MAINTAIN
;    TITLE OF INVENTION:  MALE STERILE PLANTS
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/485,139
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  010830-096
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4824 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  plasmid pCOL9
;    FEATURE:
;      NAME/KEY:  -
;      LOCATION:  396..401
;      OTHER INFORMATION:  /label= EcoRI
;    FEATURE:
;      NAME/KEY:  -
;      LOCATION:  2367..2379
;      OTHER INFORMATION:  /label= SfiI
;    FEATURE:
;      NAME/KEY:  -
;      LOCATION:  884..888
;      OTHER INFORMATION:  /label= C1-S
;      OTHER INFORMATION:  /note= "TGCAG (in C1) which in C1-S allele is
;      OTHER INFORMATION:  replaced with TTAGG"
US-08-485-139-5

  Query Match           35.1%;  Score 351.8;  DB 2;  Length 4824;
  Best Local Similarity  88.6%;  Pred. No. 3.6e-98;
  Matches  356;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy        600 ATCCAAATGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCCTGTGTGAA 659
              ||  | || |                                          ||||||||||
Db       2566 ATATATATAGTCATGCGTACAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAA 2625

Qy        660 ATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCT 719
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2626 ATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCT 2685

Qy        720 GGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCC 779
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2686 GGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCC 2745

Qy        780 AGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCG 839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2746 AGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCG 2805

Qy        840 GTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTC 899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       2806 GTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTC 2865

Qy        900 GGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAG 959
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2866 GGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAG 2925

Qy        960 GGGGATACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1001
              |||   ||||||||||||||||||||||||||||||||||||
Db       2926 GGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 2967


RESULT 7
US-08-750-357-5
; Sequence 5, Application US/08750357
; Patent No. 6008437
;  GENERAL INFORMATION:
;    APPLICANT:  KREBBERS, Enno
;    APPLICANT:  WILLIAMS, Mark
;    APPLICANT:  LEEMANS, Jan
;    TITLE OF INVENTION:  USE OF ANTHOCYANIN GENES TO MAINTAIN
;    TITLE OF INVENTION:  MALE STERILE PLANTS
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BURNS, DOANE, SWECKER & MATHIS, L.L.P.
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/750,357
;      FILING DATE:  21-MAR-1997
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  018030-100
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4824 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  plasmid pCOL9
;    FEATURE:
;      NAME/KEY:  -
;      LOCATION:  396..401
;      OTHER INFORMATION:  /label= EcoRI
;    FEATURE:
;      NAME/KEY:  -
;      LOCATION:  2367..2379
;      OTHER INFORMATION:  /label= SfiI
;    FEATURE:
;      NAME/KEY:  -
;      LOCATION:  884..888
;      OTHER INFORMATION:  /label= C1-S
;      OTHER INFORMATION:  /note= "TGCAG (in C1) which in C1-S allele is replaced with
;      OTHER INFORMATION:  TTAGG"
US-08-750-357-5

  Query Match           35.1%;  Score 351.8;  DB 3;  Length 4824;
  Best Local Similarity  88.6%;  Pred. No. 3.6e-98;
  Matches  356;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy        600 ATCCAAATGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCCTGTGTGAA 659
              ||  |  ||  |                                            ||||||||||
Db       2566 ATATATATAGTCATGCGTACAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAA 2625

Qy        660 ATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCT 719
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2626 ATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCT 2685

Qy        720 GGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCC 779
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2686 GGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCC 2745

Qy        780 AGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCG 839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2746 AGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCG 2805

Qy        840 GTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTC 899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2806 GTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTC 2865

Qy        900 GGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAG 959
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2866 GGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAG 2925

Qy        960 GGGGATACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1001
              |||   ||||||||||||||||||||||||||||||||||||
Db       2926 GGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 2967


RESULT 8
US-08-480-882B-3/c
; Sequence 3, Application US/08480882B
; Patent No. 5656275
;  GENERAL INFORMATION:
;    APPLICANT:  WASMOEN, TERRI
;    APPLICANT:  CHAVEZ, LLOYD
;    APPLICANT:  CHU, HSIEN-JUE
;    TITLE OF INVENTION:  RECOMBINANT RACCOON POX VIRUSES
;    TITLE OF INVENTION:  AND THEIR USE AS AN EFFECTIVE VACCINE AGAINST FELINE
;    TITLE OF INVENTION:  INFECTIOUS PERITONITIS VIRUS DISEASE
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Darby & Darby, P.C.
;      STREET:  805 Third Avenue  27th Floor
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,882B
;      FILING DATE:  07-JUN-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/125,516
;      FILING DATE:  22-SEP-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Green, Reza
;      REGISTRATION NUMBER:  38,475
;      REFERENCE/DOCKET NUMBER:  0632/08669
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  2125277700
;      TELEFAX:  2127536237
;      TELEX:  236687
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8710 base pairs
;      TYPE:  nucleic acid
```

```
;       STRANDEDNESS:  double
;       TOPOLOGY:  circular
;     MOLECULE TYPE:  DNA (genomic)
;     ORIGINAL SOURCE:
;       ORGANISM:  Feline infectious peritonitis virus
;     IMMEDIATE SOURCE:
;       CLONE:  psc11f1
US-08-480-882B-3

  Query Match             35.1%;  Score 351.8;  DB 1;  Length 8710;
  Best Local Similarity   88.6%;  Pred. No. 4.8e-98;
  Matches  356;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy        600 ATCCAAATGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCCTGTGTGAA 659
              |||||  |                                              ||||||||||
Db       2257 ATCCATTTATTGATCGCAAAAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAA 2198

Qy        660 ATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCT 719
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2197 ATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCT 2138

Qy        720 GGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCC 779
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2137 GGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCC 2078

Qy        780 AGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCG 839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2077 AGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCG 2018

Qy        840 GTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTC 899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2017 GTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTC 1958

Qy        900 GGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAG 959
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1957 GGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAG 1898

Qy        960 GGGGATACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1001
              |||   |||||||||||||||||||||||||||||||||||
Db       1897 GGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1856


RESULT 9
US-08-480-210-3/c
; Sequence 3, Application US/08480210
; Patent No. 5770211
;  GENERAL INFORMATION:
;    APPLICANT:  WASMOEN, TERRI
;    APPLICANT:  CHAVEZ, LLOYD
;    APPLICANT:  CHU, HSIEN-JUE
;    TITLE OF INVENTION:  RECOMBINANT RACCOON POX VIRUSES
;    TITLE OF INVENTION:  AND THEIR USE AS AN EFFECTIVE VACCINE AGAINST FELINE
;    TITLE OF INVENTION:  INFECTIOUS PERITONITIS VIRUS DISEASE
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Darby & Darby, P.C.
;      STREET:  805 Third Avenue  27th Floor
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,210
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/125,516
;      FILING DATE:  22-SEP-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Green, Reza
;      REGISTRATION NUMBER:  38,475
;      REFERENCE/DOCKET NUMBER:  0632/18669-US1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  2125277700
;      TELEFAX:  2127536237
;      TELEX:  236687
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8710 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Feline infectious peritonitis virus
;    IMMEDIATE SOURCE:
;      CLONE:  psc11f1
US-08-480-210-3

  Query Match             35.1%;  Score 351.8;  DB 1;  Length 8710;
  Best Local Similarity   88.6%;  Pred. No. 4.8e-98;
  Matches  356;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy        600 ATCCAAATGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCCTGTGTGAA 659
              |||||  |                                              ||||||||||
Db       2257 ATCCATTTATTGATCGCAAAAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAA 2198

Qy        660 ATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCT 719
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2197 ATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCT 2138

Qy        720 GGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCC 779
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2137 GGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCC 2078

Qy        780 AGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCG 839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2077 AGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCG 2018

Qy        840 GTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTC 899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2017 GTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTC 1958

Qy        900 GGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAG 959
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1957 GGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAG 1898

Qy        960 GGGGATACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1001
              |||   |||||||||||||||||||||||||||||||||||
Db       1897 GGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1856


RESULT 10
US-08-480-882B-4/c
; Sequence 4, Application US/08480882B
; Patent No. 5656275
;  GENERAL INFORMATION:
;    APPLICANT:  WASMOEN, TERRI
;    APPLICANT:  CHAVEZ, LLOYD
;    APPLICANT:  CHU, HSIEN-JUE
;    TITLE OF INVENTION:  RECOMBINANT RACCOON POX VIRUSES
;    TITLE OF INVENTION:  AND THEIR USE AS AN EFFECTIVE VACCINE AGAINST FELINE
;    TITLE OF INVENTION:  INFECTIOUS PERITONITIS VIRUS DISEASE
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Darby & Darby, P.C.
;      STREET:  805 Third Avenue  27th Floor
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10022
```

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,882B
;      FILING DATE:  07-JUN-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/125,516
;      FILING DATE:  22-SEP-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Green, Reza
;      REGISTRATION NUMBER:  38,475
;      REFERENCE/DOCKET NUMBER:  0632/08669
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  2125277700
;      TELEFAX:  2127536237
;      TELEX:  236687
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9019 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Feline immunodeficiency virus
;    IMMEDIATE SOURCE:
;      CLONE:  psc11e1
US-08-480-882B-4

  Query Match           35.1%;  Score 351.8;  DB 1;  Length 9019;
  Best Local Similarity   88.6%;  Pred. No. 4.9e-98;
  Matches  356;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy        600 ATCCAAATGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCCTGTGTGAA 659
              |||||  |                                            ||||||||||
Db       2256 ATCCATTTATTGATCGCAAAAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAA 2197

Qy        660 ATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCT 719
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2196 ATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCT 2137

Qy        720 GGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCC 779
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2136 GGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCC 2077

Qy        780 AGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCG 839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2076 AGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCG 2017

Qy        840 GTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTC 899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2016 GTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTC 1957

Qy        900 GGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAG 959
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1956 GGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAG 1897

Qy        960 GGGGATACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1001
              |||   ||||||||||||||||||||||||||||||||||||
Db       1896 GGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1855


RESULT 11
US-08-480-210-4/c
; Sequence 4, Application US/08480210
; Patent No. 5770211
;  GENERAL INFORMATION:
;    APPLICANT:  WASMOEN, TERRI
;    APPLICANT:  CHAVEZ, LLOYD
;    APPLICANT:  CHU, HSIEN-JUE
;    TITLE OF INVENTION:  RECOMBINANT RACCOON POX VIRUSES
;    TITLE OF INVENTION:  AND THEIR USE AS AN EFFECTIVE VACCINE AGAINST FELINE
;    TITLE OF INVENTION:  INFECTIOUS PERITONITIS VIRUS DISEASE
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Darby & Darby, P.C.
;      STREET:  805 Third Avenue  27th Floor
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,210
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/125,516
;      FILING DATE:  22-SEP-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Green, Reza
;      REGISTRATION NUMBER:  38,475
;      REFERENCE/DOCKET NUMBER:  0632/18669-US1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  2125277700
;      TELEFAX:  2127536237
;      TELEX:  236687
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9019 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Feline immunodeficiency virus
;    IMMEDIATE SOURCE:
;      CLONE:  psc11e1
US-08-480-210-4

  Query Match           35.1%;  Score 351.8;  DB 1;  Length 9019;
  Best Local Similarity   88.6%;  Pred. No. 4.9e-98;
  Matches  356;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy        600 ATCCAAATGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCCTGTGTGAA 659
              |||||  |                                            ||||||||||
Db       2256 ATCCATTTATTGATCGCAAAAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAA 2197

Qy        660 ATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCT 719
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2196 ATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCT 2137

Qy        720 GGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCC 779
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2136 GGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCC 2077

Qy        780 AGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCG 839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2076 AGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCG 2017

Qy        840 GTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTC 899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2016 GTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTC 1957

Qy        900 GGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAG 959
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1956 GGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAG 1897
```

```
Qy        960 GGGGATACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1001
              |||   |||||||||||||||||||||||||||||||||||
Db       1896 GGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1855


RESULT 12
US-08-726-462B-3
; Sequence 3, Application US/08726462B
; Patent No. 5800996
;  GENERAL INFORMATION:
;    APPLICANT:  Perkin-Elmer Corporation, Applied Biosystems
;    APPLICANT:  Division
;    TITLE OF INVENTION:  ENERGY TRANSFER DYES WITH ENHANCED
;    TITLE OF INVENTION:  FLUORESCENCE
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  David J. Weitz, Wilson Sonsini Goodrich
;      ADDRESSEE:  & Rosati
;      STREET:  650 Page Mill Road
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94304-1050
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch diskette
;      COMPUTER:  IBM compatible
;      OPERATING SYSTEM:  Microsoft Windows 3.1/DOS 5.0
;      SOFTWARE:  Wordperfect for windows 6.0,
;      SOFTWARE:  ASCII (DOS) TEXT format
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/726,462B
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/642,330
;      FILING DATE:  May 3, 1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/672,196
;      FILING DATE:  June 27, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  David J. Weitz
;      REGISTRATION NUMBER:  38,362
;      REFERENCE/DOCKET NUMBER:  PELM4304
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 493-9300
;      TELEFAX:  (415) 493-6811
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  738 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-726-462B-3

  Query Match             35.0%;  Score 350.8;  DB 1;  Length 738;
  Best Local Similarity   88.5%;  Pred. No. 2.9e-98;
  Matches  355;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy        601 TCCAAATGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCCTGTGTGAAA 660
              || | |  |                                        ||||||||||||
Db         83 TCTATAGTGTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAA 142

Qy        661 TTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 TTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTG 202

Qy        721 GGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        203 GGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCA 262

Qy        781 GTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        263 GTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGG 322

Qy        841 TTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        323 TTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCG 382

Qy        901 GCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        383 GCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGG 442

Qy        961 GGGATACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1001
              ||   |||||||||||||||||||||||||||||||||||
Db        443 GGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 483


RESULT 13
US-09-046-203-3
; Sequence 3, Application US/09046203
; Patent No. 5945526
;  GENERAL INFORMATION:
;    APPLICANT:  Perkin-Elmer Corporation, Applied Biosystems
;    APPLICANT:  Division
;    TITLE OF INVENTION:  ENERGY TRANSFER DYES WITH ENHANCED
;    TITLE OF INVENTION:  FLUORESCENCE
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  David J. Weitz, Wilson Sonsini Goodrich
;      ADDRESSEE:  & Rosati
;      STREET:  650 Page Mill Road
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94304-1050
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch diskette
;      COMPUTER:  IBM compatible
;      OPERATING SYSTEM:  Microsoft Windows 3.1/DOS 5.0
;      SOFTWARE:  Wordperfect for windows 6.0,
;      SOFTWARE:  ASCII (DOS) TEXT format
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/046,203
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/642,330
;      FILING DATE:  May 3, 1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/672,196
;      FILING DATE:  June 27, 1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/726,462
;      FILING DATE:  October 4, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  David J. Weitz
;      REGISTRATION NUMBER:  38,362
;      REFERENCE/DOCKET NUMBER:  16842-746
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 493-9300
;      TELEFAX:  (415) 493-6811
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  738 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-046-203-3

  Query Match             35.0%;  Score 350.8;  DB 2;  Length 738;
  Best Local Similarity   88.5%;  Pred. No. 2.9e-98;
  Matches  355;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy        601 TCCAAATGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCCTGTGTGAAA 660
              || | |  |                                        ||||||||||||
Db         83 TCTATAGTGTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAA 142
```

```
Qy        661 TTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 TTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTG 202

Qy        721 GGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        203 GGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCA 262

Qy        781 GTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        263 GTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGG 322

Qy        841 TTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        323 TTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCG 382

Qy        901 GCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        383 GCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGG 442

Qy        961 GGGATACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1001
              ||  ||||||||||||||||||||||||||||||||||||
Db        443 GGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 483


RESULT 14
US-09-272-104-3
; Sequence 3, Application US/09272104
; Patent No. 6145434
;  GENERAL INFORMATION:
;    APPLICANT:  Perkin-Elmer Corporation, Applied Biosystems
;    APPLICANT:  Division
;    TITLE OF INVENTION:  METHOD FOR DETECTING
;    TITLE OF INVENTION:  OLIGONUCLEOTIDES USING ENERGY TRANSFER DYES WITH LONG STOKE SH
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  David J. Weitz, Wilson Sonsini Goodrich
;      ADDRESSEE:  & Rosati
;      STREET:  650 Page Mill Road
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94304-1050
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch diskette
;      COMPUTER:  IBM compatible
;      OPERATING SYSTEM:  Microsoft Windows 3.1/DOS 5.0
;      SOFTWARE:  Wordperfect for windows 6.0,
;      SOFTWARE:  ASCII (DOS) TEXT format
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/272,104
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/642,330
;      FILING DATE:  May 3, 1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/672,196
;      FILING DATE:  June 27, 1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/726,462
;      FILING DATE:  October 4, 1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  09/046,203
;      FILING DATE:  March 23, 1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  David J. Weitz
;      REGISTRATION NUMBER:  38,362
;      REFERENCE/DOCKET NUMBER:  16842-755
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 493-9300
;      TELEFAX:  (415) 493-6811

;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  738 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-272-104-3

  Query Match             35.0%;  Score 350.8;  DB 3;  Length 738;
  Best Local Similarity   88.5%;  Pred. No. 2.9e-98;
  Matches  355;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy        601 TCCAAATGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCCTGTGTGAAA 660
              || | |  |                                           ||||||||||||
Db         83 TCTATAGTGTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAA 142

Qy        661 TTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 TTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTG 202

Qy        721 GGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        203 GGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCA 262

Qy        781 GTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        263 GTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGG 322

Qy        841 TTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        323 TTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCG 382

Qy        901 GCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        383 GCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGG 442

Qy        961 GGGATACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1001
              ||  ||||||||||||||||||||||||||||||||||||
Db        443 GGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 483


RESULT 15
US-09-272-097-3
; Sequence 3, Application US/09272097
; Patent No. 6335440
;   GENERAL INFORMATION:
;        APPLICANT: PE Corporation (NY)
;        TITLE OF INVENTION: METHOD FOR DETECTING OLIGONUCLEOTIDES USING
;                            ENERGY TRANSFER DYES WITH LONG STOKE SHIFT
;        NUMBER OF SEQUENCES: 3
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: David J. Weitz, Wilson Sonsini Goodrich
;                        & Rosati
;             STREET: 650 Page Mill Road
;             CITY: Palo Alto
;             STATE: California
;             COUNTRY: USA
;             ZIP: 94304-1050
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3.5 inch diskette
;             COMPUTER: IBM compatible
;             OPERATING SYSTEM: Microsoft Windows 3.1/DOS 5.0
;             SOFTWARE: Word 97
;                       ASCII (DOS) TEXT format
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/272,097
;             FILING DATE: 18-Mar-1999
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/642,330
;             FILING DATE: May 3, 1996
;             APPLICATION NUMBER: 08/672,196
;             FILING DATE: June 27, 1996
```

; APPLICATION NUMBER: 08/726,462
; FILING DATE: October 4, 1996
; APPLICATION NUMBER: 09/046,203
; FILING DATE: March 23, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: David J. Weitz
; REGISTRATION NUMBER: 38,362
; REFERENCE/DOCKET NUMBER: 16842-755
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 493-9300
; TELEFAX: (650) 493-6811
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 738 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-272-097-3

Query Match 35.0%; Score 350.8; DB 4; Length 738;
Best Local Similarity 88.5%; Pred. No. 2.9e-98;
Matches 355; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

```
Qy        601 TCCAAATGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCCTGTGTGAAA 660
              || | |  |                                       ||||||||||||
Db         83 TCTATAGTGTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAA 142

Qy        661 TTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 TTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTG 202

Qy        721 GGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        203 GGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCA 262

Qy        781 GTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        263 GTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGG 322

Qy        841 TTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        323 TTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCG 382

Qy        901 GCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        383 GCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGG 442

Qy        961 GGGATACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1001
              ||   ||||||||||||||||||||||||||||||||||||
Db        443 GGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 483
```

Search completed: July 16, 2003, 14:58:48
Job time : 42.0516 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 16, 2003, 13:28:07 ; Search time 131.716 Seconds
(without alignments)
12133.023 Million cell updates/sec

Title: US-09-981-353-24
Perfect score: 1001
Sequence: 1 atgattacgaattcgagctc.....gtgagcaaaaggccagcaaa 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1114047 seqs, 798260406 residues

Total number of hits satisfying chosen parameters: 2228094

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
6: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	962	96.1	1001	9	US-09-981-353-24	Sequence 24, Appl
	2	583	58.2	737	9	US-10-235-994-17	Sequence 17, Appl
	3	582	58.1	736	9	US-10-158-646-5	Sequence 5, Appli
c	4	352.8	35.2	7208	9	US-10-229-346-5	Sequence 5, Appli
	5	352.4	35.2	4069	9	US-10-251-385-287	Sequence 287, App
c	6	352.4	35.2	4069	9	US-10-251-385-288	Sequence 288, App
	7	352	35.2	3983	10	US-09-758-987-1	Sequence 1, Appli
c	8	351.8	35.1	8108	9	US-10-121-988-77	Sequence 77, Appl
	9	350.8	35.0	584	9	US-09-736-457-639	Sequence 639, App
	10	350.8	35.0	584	9	US-09-902-941-639	Sequence 639, App
	11	350.8	35.0	584	9	US-09-849-626-639	Sequence 639, App
	12	350.8	35.0	584	9	US-10-017-754-639	Sequence 639, App
	13	350.8	35.0	584	9	US-09-476-300-639	Sequence 639, App
c	14	350.8	35.0	2571	9	US-10-137-036-93	Sequence 93, Appl
	15	350.8	35.0	3216	9	US-10-057-108-1	Sequence 1, Appli
	16	350.8	35.0	3357	9	US-10-161-403-96	Sequence 96, Appl
	17	350.8	35.0	3448	9	US-10-128-590-6	Sequence 6, Appli
	18	350.8	35.0	4118	10	US-09-068-821-17	Sequence 17, Appl
	19	350.8	35.0	4118	10	US-09-068-821-18	Sequence 18, Appl
	20	350.8	35.0	4162	10	US-09-861-881-1	Sequence 1, Appli
	21	350.8	35.0	4393	10	US-09-861-881-2	Sequence 2, Appli
	22	350.8	35.0	6688	9	US-09-813-453A-72	Sequence 72, Appl
	23	350.8	35.0	8320	9	US-09-813-453A-71	Sequence 71, Appl
c	24	350.6	35.0	4689	9	US-10-057-108-5	Sequence 5, Appli
c	25	350.6	35.0	6470	9	US-10-057-108-4	Sequence 4, Appli
c	26	350.6	35.0	8330	9	US-10-057-108-3	Sequence 3, Appli
	27	350.2	35.0	752	10	US-09-956-004-108	Sequence 108, App
c	28	350.2	35.0	1092	9	US-09-764-868-232	Sequence 232, App
c	29	350.2	35.0	1092	9	US-09-764-891-1853	Sequence 1853, Ap
c	30	350.2	35.0	2074	9	US-10-106-698-350	Sequence 350, App
	31	350.2	35.0	2192	9	US-10-021-403A-10	Sequence 10, Appl
	32	350.2	35.0	3534	9	US-10-021-403A-9	Sequence 9, Appli
	33	350.2	35.0	3589	9	US-10-234-406-1	Sequence 1, Appli
	34	350.2	35.0	3609	9	US-10-234-406-3	Sequence 3, Appli
c	35	350.2	35.0	3637	9	US-10-206-030-3	Sequence 3, Appli
c	36	350.2	35.0	3637	9	US-10-206-030-4	Sequence 4, Appli
c	37	350.2	35.0	3637	12	US-10-066-390-3	Sequence 3, Appli
c	38	350.2	35.0	3637	12	US-10-066-390-4	Sequence 4, Appli
	39	350.2	35.0	3928	9	US-10-128-590-45	Sequence 45, Appl
	40	350.2	35.0	4119	10	US-09-993-170-1	Sequence 1, Appli
	41	350.2	35.0	4205	9	US-09-883-848A-35	Sequence 35, Appl
	42	350.2	35.0	4205	9	US-09-883-848A-48	Sequence 48, Appl
	43	350.2	35.0	4496	9	US-10-234-406-5	Sequence 5, Appli
	44	350.2	35.0	4704	9	US-09-932-328-4	Sequence 4, Appli
c	45	350.2	35.0	4773	9	US-09-991-209-32	Sequence 32, Appl

ALIGNMENTS

RESULT 1
US-09-981-353-24
; Sequence 24, Application US/09981353
; Patent No. US20020160382A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy W.
; APPLICANT: Jones, David A.
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
; FILE REFERENCE: PA-0038 US
; CURRENT APPLICATION NUMBER: US/09/981,353
; CURRENT FILING DATE: 2001-10-11
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 24
; LENGTH: 1001
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020160382A1 3220207CB1
; NAME/KEY: unsure
; LOCATION: 610-648
; OTHER INFORMATION: a, t, c, g, or other
US-09-981-353-24

Query Match 96.1%; Score 962; DB 9; Length 1001;
Best Local Similarity 100.0%; Pred. No. 3e-309;
Matches 1001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGATTACGAATTCGAGCTCGTACCCCCCCAGAATGTTGACAGTCGCTCTCCTAGCCCTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGATTACGAATTCGAGCTCGTACCCCCCCAGAATGTTGACAGTCGCTCTCCTAGCCCTT 60

Qy         61 CTCTGTGCCTCAGCCTCTGGCAATGCCATTCAGGCCAGGTCTTCCTCCTATAGTGGAGAG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CTCTGTGCCTCAGCCTCTGGCAATGCCATTCAGGCCAGGTCTTCCTCCTATAGTGGAGAG 120

Qy        121 TATGGAAGTGGTGGTGGAAAGCGATTCTCTCATTCTGGCAACCAGTTGGACGGCCCCATC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TATGGAAGTGGTGGTGGAAAGCGATTCTCTCATTCTGGCAACCAGTTGGACGGCCCCATC 180

Qy        181 ACCGCCCTCCGGGTCCGAGTCAACACATACTACATCGTAGGTCTTCAGGTGCGCTATGGC 240
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACCGCCCTCCGGGTCCGAGTCAACACATACTACATCGTAGGTCTTCAGGTGCGCTATGGC 240

Qy        241 AAGGTGTGGAGCGACTATGTGGGTGGTCGCAACGGAGACCTGGAGGAGATCTTTCTGCAC 300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGGTGTGGAGCGACTATGTGGGTGGTCGCAACGGAGACCTGGAGGAGATCTTTCTGCAC 300

Qy        301 CCTGGGGAATCAGTGATCCAGGTTTCTGGGAAGTACAAGTGGTACCTGAAGAAGCTGGTA 360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CCTGGGGAATCAGTGATCCAGGTTTCTGGGAAGTACAAGTGGTACCTGAAGAAGCTGGTA 360

Qy        361 TTTGTGACAGACAAGGGCCGCTATCTGTCTTTTGGGAAAGACAGTGGCACAAGTTTCAAT 420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TTTGTGACAGACAAGGGCCGCTATCTGTCTTTTGGGAAAGACAGTGGCACAAGTTTCAAT 420

Qy        421 GCCGTCCCCTTGCACCCCAACACCGTGCTCCGCTTCATCAGTGGCCGGTCTGGTTCTCTC 480
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GCCGTCCCCTTGCACCCCAACACCGTGCTCCGCTTCATCAGTGGCCGGTCTGGTTCTCTC 480

Qy        481 ATCGATGCCATTGGCCTGCACTGGGATGTTTACCCCACTAGCTGCAGCAGATGCTGAGCC 540
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATCGATGCCATTGGCCTGCACTGGGATGTTTACCCCACTAGCTGCAGCAGATGCTGAGCC 540

Qy        541 TCCTCTCCTTGGCAGGGGCACTGTGATGAGGAGTAAGAACTCCCTTATCACTAACCCCCA 600
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TCCTCTCCTTGGCAGGGGCACTGTGATGAGGAGTAAGAACTCCCTTATCACTAACCCCCA 600

Qy        601 TCCAAATGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCCTGTGTGAAA 660
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TCCAAATGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCCTGTGTGAAA 660

Qy        661 TTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTG 720
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTG 720

Qy        721 GGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCA 780
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCA 780

Qy        781 GTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGG 840
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGG 840

Qy        841 TTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCG 900
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCG 900

Qy        901 GCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGG 960
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGG 960

Qy        961 GGGATACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1001
                |||||||||||||||||||||||||||||||||||||||||
Db        961 GGGATACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1001


RESULT 2
US-10-235-994-17
; Sequence 17, Application US/10235994
; Publication No. US20030101002A1
; GENERAL INFORMATION:
;  APPLICANT: Bartha, Gabor
;  APPLICANT:  Walker, Michael
;  TITLE OF INVENTION: METHODS FOR ANALYZING GENE EXPRESSION PATTERNS
;  FILE REFERENCE: ICYTP012
;  CURRENT APPLICATION NUMBER: US/10/235,994
;  CURRENT FILING DATE:  2002-09-04
;  PRIOR APPLICATION NUMBER: US/10/003,608
;  PRIOR FILING DATE: 2001-11-01
;  PRIOR APPLICATION NUMBER: 60/245,081
;  PRIOR FILING DATE: 2000-11-01
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 737
;   TYPE: DNA
;   ORGANISM: Human
US-10-235-994-17

  Query Match            58.2%;  Score 583;  DB 9;  Length 737;
  Best Local Similarity  100.0%;  Pred. No. 2.7e-183;
  Matches  583;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 CCCCAGAATGTTGACAGTCGCTCTCCTAGCCCTTCTCTGTGCCTCAGCCTCTGGCAATGC 86
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        129 CCCCAGAATGTTGACAGTCGCTCTCCTAGCCCTTCTCTGTGCCTCAGCCTCTGGCAATGC 188

Qy         87 CATTCAGGCCAGGTCTTCCTCCTATAGTGGAGAGTATGGAAGTGGTGGTGGAAAGCGATT 146
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        189 CATTCAGGCCAGGTCTTCCTCCTATAGTGGAGAGTATGGAAGTGGTGGTGGAAAGCGATT 248

Qy        147 CTCTCATTCTGGCAACCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACAC 206
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        249 CTCTCATTCTGGCAACCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACAC 308

Qy        207 ATACTACATCGTAGGTCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGG 266
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        309 ATACTACATCGTAGGTCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGG 368

Qy        267 TCGCAACGGAGACCTGGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTC 326
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        369 TCGCAACGGAGACCTGGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTC 428

Qy        327 TGGGAAGTACAAGTGGTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCT 386
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        429 TGGGAAGTACAAGTGGTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCT 488

Qy        387 GTCTTTTGGGAAAGACAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGT 446
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        489 GTCTTTTGGGAAAGACAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGT 548

Qy        447 GCTCCGCTTCATCAGTGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGA 506
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        549 GCTCCGCTTCATCAGTGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGA 608

Qy        507 TGTTTACCCCACTAGCTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGA 566
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        609 TGTTTACCCCACTAGCTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGA 668

Qy        567 TGAGGAGTAAGAACTCCCTTATCACTAACCCCCATCCAAATGG 609
               |||||||||||||||||||||||||||||||||||||||||||
Db        669 TGAGGAGTAAGAACTCCCTTATCACTAACCCCCATCCAAATGG 711


RESULT 3
US-10-158-646-5
; Sequence 5, Application US/10158646
; Publication No. US20030073105A1
; GENERAL INFORMATION:
;  APPLICANT: Lasek, Amy K.W.
;  APPLICANT:  Sornasse, Thierry
;  TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
;  FILE REFERENCE: PA-0030-1 US
;  CURRENT APPLICATION NUMBER: US/10/158,646
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: 60/295,239
;  PRIOR FILING DATE: 2001-05-31
;  NUMBER OF SEQ ID NOS: 78
;  SOFTWARE: PERL Program
; SEQ ID NO 5
;   LENGTH: 736
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20030073105A1 239588.4
```

US-10-158-646-5

Query Match 58.1%; Score 582; DB 9; Length 736;
Best Local Similarity 100.0%; Pred. No. 5.8e-183;
Matches 582; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         27 CCCCAGAATGTTGACAGTCGCTCTCCTAGCCCTTCTCTGTGCCTCAGCCTCTGGCAATGC 86
Db        129 CCCCAGAATGTTGACAGTCGCTCTCCTAGCCCTTCTCTGTGCCTCAGCCTCTGGCAATGC 188

Qy         87 CATTCAGGCCAGGTCTTCCTCCTATAGTGGAGAGTATGGAAGTGGTGGTGGAAAGCGATT 146
Db        189 CATTCAGGCCAGGTCTTCCTCCTATAGTGGAGAGTATGGAAGTGGTGGTGGAAAGCGATT 248

Qy        147 CTCTCATTCTGGCAACCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACAC 206
Db        249 CTCTCATTCTGGCAACCAGTTGGACGGCCCCATCACCGCCCTCCGGGTCCGAGTCAACAC 308

Qy        207 ATACTACATCGTAGGTCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGG 266
Db        309 ATACTACATCGTAGGTCTTCAGGTGCGCTATGGCAAGGTGTGGAGCGACTATGTGGGTGG 368

Qy        267 TCGCAACGGAGACCTGGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTC 326
Db        369 TCGCAACGGAGACCTGGAGGAGATCTTTCTGCACCCTGGGGAATCAGTGATCCAGGTTTC 428

Qy        327 TGGGAAGTACAAGTGGTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCT 386
Db        429 TGGGAAGTACAAGTGGTACCTGAAGAAGCTGGTATTTGTGACAGACAAGGGCCGCTATCT 488

Qy        387 GTCTTTTGGGAAAGACAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGT 446
Db        489 GTCTTTTGGGAAAGACAGTGGCACAAGTTTCAATGCCGTCCCCTTGCACCCCAACACCGT 548

Qy        447 GCTCCGCTTCATCAGTGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGA 506
Db        549 GCTCCGCTTCATCAGTGGCCGGTCTGGTTCTCTCATCGATGCCATTGGCCTGCACTGGGA 608

Qy        507 TGTTTACCCCACTAGCTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGA 566
Db        609 TGTTTACCCCACTAGCTGCAGCAGATGCTGAGCCTCCTCTCCTTGGCAGGGGCACTGTGA 668

Qy        567 TGAGGAGTAAGAACTCCCTTATCACTAACCCCCATCCAAATG 608
Db        669 TGAGGAGTAAGAACTCCCTTATCACTAACCCCCATCCAAATG 710
```

RESULT 4
US-10-229-346-5/c
; Sequence 5, Application US/10229346
; Publication No. US20030120054A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Eric
; APPLICANT: Stacy, Cheryl
; TITLE OF INVENTION: Modified Cry3A Toxins
; FILE REFERENCE: 60065A
; CURRENT APPLICATION NUMBER: US/10/229,346
; CURRENT FILING DATE: 2002-08-27
; PRIOR APPLICATION NUMBER: 60/316,421
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 7208
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: pCIB6850
US-10-229-346-5

Query Match 35.2%; Score 352.8; DB 9; Length 7208;
Best Local Similarity 82.4%; Pred. No. 2.4e-106;
Matches 378; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

```
Qy        543 CTCTCCTTGGCAGGGGCACTGTGATGAGGAGTAAGAACTCCCTTATCACTAACCCCCATC 602
Db       4727 CTATACTTCACAGCCCGTCGCAGAAGTATTGTTGGATCGTTGTGGTCTCCATCCTCTTAC 4668

Qy        603 CAAATGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCCTGTGTGAAATT 662
Db       4667 AATTGTTGTCATGTGCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATT 4608

Qy        663 GTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGG 722
Db       4607 GTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGG 4548

Qy        723 GTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGT 782
Db       4547 GTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGT 4488

Qy        783 CGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTT 842
Db       4487 CGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTT 4428

Qy        843 TGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGC 902
Db       4427 TGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGC 4368

Qy        903 TGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGG 962
Db       4367 TGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGG 4308

Qy        963 GATACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1001
Db       4307 ATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 4269
```

RESULT 5
US-10-251-385-287
; Sequence 287, Application US/10251385
; Publication No. US20030105292A1
; GENERAL INFORMATION:
; APPLICANT: Behan, Dominic P.
; APPLICANT: Chalmers, Derek T.
; APPLICANT: Liaw, Chen W.
; TITLE OF INVENTION: No. US20030105292A1-Endogenous, Constitutively Activated Human
; TITLE OF INVENTION: Protein-Coupled
; TITLE OF INVENTION: Receptors
; FILE REFERENCE: AREN-0040
; CURRENT APPLICATION NUMBER: US/10/251,385
; CURRENT FILING DATE: 2002-09-20
; PRIOR APPLICATION NUMBER: US/09/170,496
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 294
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 287
; LENGTH: 4069
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-251-385-287

Query Match 35.2%; Score 352.4; DB 9; Length 4069;
Best Local Similarity 88.8%; Pred. No. 2.6e-106;
Matches 356; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

```
Qy        601 TCCAAATGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCCTGTGTGAAA 660
Db       1299 TGCAAAAAGCTCCCTCGAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAA 1358

Qy        661 TTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTG 720
Db       1359 TTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTG 1418

Qy        721 GGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCA 780
```

```
Db       1419 GGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCA 1478

Qy        781 GTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1479 GTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGG 1538

Qy        841 TTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1539 TTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCG 1598

Qy        901 GCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1599 GCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGG 1658

Qy        961 GGGATACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1001
              || ||||||||||||||||||||||||||||||||||||
Db       1659 GGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1699


RESULT 6
US-10-251-385-288/c
; Sequence 288, Application US/10251385
; Publication No. US20030105292A1
; GENERAL INFORMATION:
;  APPLICANT: Behan, Dominic P.
;  APPLICANT:  Chalmers, Derek T.
;  APPLICANT:  Liaw, Chen W.
;  TITLE OF INVENTION: No. US20030105292A1-Endogenous, Constitutively Activated Human G
;  TITLE OF INVENTION:  Protein-Coupled
;  TITLE OF INVENTION:  Receptors
;  FILE REFERENCE: AREN-0040
;  CURRENT APPLICATION NUMBER: US/10/251,385
;  CURRENT FILING DATE:  2002-09-20
;  PRIOR APPLICATION NUMBER: US/09/170,496
;  PRIOR FILING DATE: 1998-10-13
;  NUMBER OF SEQ ID NOS: 294
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 288
;   LENGTH: 4069
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-251-385-288

  Query Match             35.2%;  Score 352.4;  DB 9;  Length 4069;
  Best Local Similarity   88.8%;  Pred. No. 2.6e-106;
  Matches  356;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy        601 TCCAAATGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCCTGTGTGAAA 660
              | ||||  |                                          ||||||||||||
Db       2771 TGCAAAAAGCTCCCTCGAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAA 2712

Qy        661 TTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2711 TTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTG 2652

Qy        721 GGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2651 GGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCA 2592

Qy        781 GTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2591 GTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGG 2532

Qy        841 TTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2531 TTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCG 2472

Qy        901 GCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2471 GCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGG 2412

Qy        961 GGGATACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1001
              ||   ||||||||||||||||||||||||||||||||||||
Db       2411 GGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 2371


RESULT 7
US-09-758-987-1
; Sequence 1, Application US/09758987
; Patent No. US20010029014A1
; GENERAL INFORMATION:
;  APPLICANT:  De Beuckeleer, Marc
;  TITLE OF INVENTION:   Methods and kits for identifying elite event GAT-ZM1 in biol
;  TITLE OF INVENTION:  samples
;  FILE REFERENCE:   514412-2025.1
;  CURRENT APPLICATION NUMBER: US/09/758,987
;  CURRENT FILING DATE:  2001-01-11
;  PRIOR APPLICATION NUMBER:   09/481,049
;  PRIOR FILING DATE:   2000-01-11
;  NUMBER OF SEQ ID NOS:   14
;  SOFTWARE:   PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 3983
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(3983)
;   OTHER INFORMATION: nucleotide sequence of the genetic elements of pUC/Ac
US-09-758-987-1

  Query Match             35.2%;  Score 352;  DB 10;  Length 3983;
  Best Local Similarity   88.0%;  Pred. No. 3.4e-106;
  Matches  358;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy        595 CCCCCATCCAAATGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCCTGT 654
              || | ||   |||                                      ||||||
Db       1720 CCTCTATTTGAATCTTTGACTCCATGGGAATTCGTAATCATGGTCATAGCTGTTTCCTGT 1779

Qy        655 GTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAA 714
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1780 GTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAA 1839

Qy        715 AGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGC 774
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1840 AGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGC 1899

Qy        775 TTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAG 834
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1900 TTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAG 1959

Qy        835 AGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGT 894
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1960 AGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGT 2019

Qy        895 CGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGA 954
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2020 CGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGA 2079

Qy        955 ATCAGGGGGATACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1001
              ||||||||  ||||||||||||||||||||||||||||||||||||
Db       2080 ATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 2126


RESULT 8
US-10-121-988-77/c
; Sequence 77, Application US/10121988
; Publication No. US20030068327A1
; GENERAL INFORMATION:
;  APPLICANT: Hosken, Nancy Ann
;  APPLICANT:  McGowan, Patrick
;  APPLICANT:  Sleath, Paul R.
;  APPLICANT:  Mossman, Sally P.
;  APPLICANT:  Evans, Lawrence S.
;  APPLICANT:  Swanson, Ryan M.
;  APPLICANT:  McNeill, Patricia D.
```

```
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
;  TITLE OF INVENTION:  TREATMENT OF HERPES SIMPLEX VIRUS INFECTION
;  FILE REFERENCE: 210121.538C1
;  CURRENT APPLICATION NUMBER: US/10/121,988
;  CURRENT FILING DATE:  2002-04-11
;  NUMBER OF SEQ ID NOS: 183
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 77
;   LENGTH: 8108
;   TYPE: DNA
;   ORGANISM: HSV-2
US-10-121-988-77

  Query Match             35.1%;  Score 351.8;  DB 9;  Length 8108;
  Best Local Similarity   88.6%;  Pred. No. 5.4e-106;
  Matches  356;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy        600 ATCCAAATGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCCTGTGTGAA 659
              |||||  |                                             ||||||||||
Db       2267 ATCCATTTATTGATCGCAAAAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAA 2208

Qy        660 ATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCT 719
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2207 ATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCT 2148

Qy        720 GGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCC 779
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2147 GGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCC 2088

Qy        780 AGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCG 839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2087 AGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCG 2028

Qy        840 GTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTC 899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2027 GTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTC 1968

Qy        900 GGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAG 959
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1967 GGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAG 1908

Qy        960 GGGGATACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1001
              |||   ||||||||||||||||||||||||||||||||||||
Db       1907 GGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1866


RESULT 9
US-09-736-457-639
; Sequence 639, Application US/09736457
; Patent No. US20020168637A1
; GENERAL INFORMATION:
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Lodes, Michael A.
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Vedvick, Tom
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Retter, Marc
;  APPLICANT:  Mannion, Jane
;  APPLICANT:  Fan, Liqun
;  APPLICANT:  Wang, Aijun
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE:  210121.478C15
;  CURRENT APPLICATION NUMBER: US/09/736,457
;  CURRENT FILING DATE:  2000-12-13
;  NUMBER OF SEQ ID NOS:  1864
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 639
;   LENGTH: 584
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-736-457-639

  Query Match             35.0%;  Score 350.8;  DB 9;  Length 584;
  Best Local Similarity   88.5%;  Pred. No. 3.8e-106;
  Matches  355;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy        601 TCCAAATGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCCTGTGTGAAA 660
              || | |  |                                             ||||||||||||
Db         11 TCTATAGTGTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAA 70

Qy        661 TTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         71 TTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTG 130

Qy        721 GGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        131 GGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCA 190

Qy        781 GTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        191 GTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGG 250

Qy        841 TTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        251 TTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCG 310

Qy        901 GCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        311 GCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGG 370

Qy        961 GGGATACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1001
              ||   ||||||||||||||||||||||||||||||||||||
Db        371 GGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 411


RESULT 10
US-09-902-941-639
; Sequence 639, Application US/09902941
; Patent No. US20020172952A1
; GENERAL INFORMATION:
;  APPLICANT: Henderson, Robert A.
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Watanabe, Yoshihiro
;  APPLICANT:  Johnson, Jeffrey C.
;  APPLICANT:  Retter, Marc W.
;  APPLICANT:  Marnerakis, Margarita
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Vedvick, Thomas S.
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  McNabb, Andria
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.478C17
;  CURRENT APPLICATION NUMBER: US/09/902,941
;  CURRENT FILING DATE:  2001-07-10
;  NUMBER OF SEQ ID NOS: 2002
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 639
;   LENGTH: 584
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-902-941-639

  Query Match             35.0%;  Score 350.8;  DB 9;  Length 584;
  Best Local Similarity   88.5%;  Pred. No. 3.8e-106;
  Matches  355;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy        601 TCCAAATGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCCTGTGTGAAA 660
              || | |  |                                             ||||||||||||
Db         11 TCTATAGTGTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAA 70

Qy        661 TTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db         71 TTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTG 130

Qy        721 GGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        131 GGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCA 190

Qy        781 GTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        191 GTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGG 250

Qy        841 TTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        251 TTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCG 310

Qy        901 GCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        311 GCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGG 370

Qy        961 GGGATACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1001
              ||   ||||||||||||||||||||||||||||||||||
Db        371 GGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 411


RESULT 11
US-09-849-626-639
; Sequence 639, Application US/09849626
; Publication No. US20020197669A1
; GENERAL INFORMATION:
;  APPLICANT: Bangur, Chaitanya
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Wang, Aijun
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Switzer, Anne
;  APPLICANT:  McNeill, Patricia
;  APPLICANT:  Clapper, Jonathan
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.478C16
;  CURRENT APPLICATION NUMBER: US/09/849,626
;  CURRENT FILING DATE:  2001-05-03
;  NUMBER OF SEQ ID NOS: 1926
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 639
;   LENGTH: 584
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-849-626-639

  Query Match             35.0%;  Score 350.8;  DB 9;  Length 584;
  Best Local Similarity   88.5%;  Pred. No. 3.8e-106;
  Matches  355;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy        601 TCCAAATGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCCTGTGTGAAA 660
              || | |  |                                          ||||||||||||
Db         11 TCTATAGTGTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAA 70

Qy        661 TTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         71 TTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTG 130

Qy        721 GGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        131 GGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCA 190

Qy        781 GTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        191 GTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGG 250

Qy        841 TTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        251 TTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCG 310

Qy        901 GCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        311 GCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGG 370

Qy        961 GGGATACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1001
              ||   ||||||||||||||||||||||||||||||||||
Db        371 GGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 411


RESULT 12
US-10-017-754-639
; Sequence 639, Application US/10017754
; Publication No. US20030054363A1
; GENERAL INFORMATION:
;  APPLICANT: Henderson, Robert A.
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Watanabe, Yoshihiro
;  APPLICANT:  Johnson, Jeffrey C.
;  APPLICANT:  Retter, Marc W.
;  APPLICANT:  Marnerakis, Margarita
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Vedvick, Thomas S.
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  McNabb, Andria
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.478C18
;  CURRENT APPLICATION NUMBER: US/10/017,754
;  CURRENT FILING DATE:  2001-10-29
;  NUMBER OF SEQ ID NOS: 2004
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 639
;   LENGTH: 584
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-017-754-639

  Query Match             35.0%;  Score 350.8;  DB 9;  Length 584;
  Best Local Similarity   88.5%;  Pred. No. 3.8e-106;
  Matches  355;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy        601 TCCAAATGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCCTGTGTGAAA 660
              || | |  |                                          ||||||||||||
Db         11 TCTATAGTGTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAA 70

Qy        661 TTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         71 TTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTG 130

Qy        721 GGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        131 GGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCA 190

Qy        781 GTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        191 GTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGG 250

Qy        841 TTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        251 TTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCG 310

Qy        901 GCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        311 GCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGG 370

Qy        961 GGGATACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1001
              ||   ||||||||||||||||||||||||||||||||||
Db        371 GGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 411


RESULT 13
US-09-476-300-639
; Sequence 639, Application US/09476300
```

```
; Publication No. US20030125245A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Tongtong
;  APPLICANT: Bangur, Chaitanya S.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
;  TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.478C3
;  CURRENT APPLICATION NUMBER: US/09/476,300
;  CURRENT FILING DATE: 1999-12-30
;  NUMBER OF SEQ ID NOS: 785
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 639
;   LENGTH: 584
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-476-300-639

  Query Match             35.0%;  Score 350.8;  DB 9;  Length 584;
  Best Local Similarity   88.5%;  Pred. No. 3.8e-106;
  Matches  355;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy        601 TCCAAATGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCCTGTGTGAAA 660
              || | |  |                                       ||||||||||||
Db         11 TCTATAGTGTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAA 70

Qy        661 TTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         71 TTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTG 130

Qy        721 GGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        131 GGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCA 190

Qy        781 GTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        191 GTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGG 250

Qy        841 TTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        251 TTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCG 310

Qy        901 GCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        311 GCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGG 370

Qy        961 GGGATACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1001
              ||   |||||||||||||||||||||||||||||||||||
Db        371 GGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 411


RESULT 14
US-10-137-036-93/c
; Sequence 93, Application US/10137036
; Publication No. US20030101478A1
; GENERAL INFORMATION:
;  APPLICANT: Perera, Ranjan
;  APPLICANT:  Rice, Stephen
;  APPLICANT:  Eagleton, Clare
;  APPLICANT:  Lasham, Annette
;  APPLICANT:  Wood, Marion
;  APPLICANT:  Visser, Elizabeth
;  TITLE OF INVENTION: Compositions and Methods for the
;  TITLE OF INVENTION:  Modification of Gene Expression
;  FILE REFERENCE: 11000.1036c4
;  CURRENT APPLICATION NUMBER: US/10/137,036
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: PCT/NZ 01/00115
;  PRIOR FILING DATE: 2001-06-20
;  PRIOR APPLICATION NUMBER: U.S. No. US20030101478A1 09/724,624
;  PRIOR FILING DATE: 2000-11-28
;  PRIOR APPLICATION NUMBER: U.S. No. US20030101478A1 09/598,401
;  PRIOR FILING DATE: 2000-06-20
;  PRIOR APPLICATION NUMBER: PCT/NZ00/00018
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: U.S. No. US20030101478A1 60/146,591
;  PRIOR FILING DATE: 1999-07-30
;  PRIOR APPLICATION NUMBER: U.S. No. US20030101478A1 09/276,599
;  PRIOR FILING DATE: 1999-03-25
;  NUMBER OF SEQ ID NOS: 143
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 93
;   LENGTH: 2571
;   TYPE: DNA
;   ORGANISM: Eucalyptus grandis
US-10-137-036-93

  Query Match             35.0%;  Score 350.8;  DB 9;  Length 2571;
  Best Local Similarity   88.5%;  Pred. No. 7.2e-106;
  Matches  355;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy        601 TCCAAATGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCCTGTGTGAAA 660
              || | |  |                                       ||||||||||||
Db        921 TCTATAGTGTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAA 862

Qy        661 TTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        861 TTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTG 802

Qy        721 GGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        801 GGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCA 742

Qy        781 GTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        741 GTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGG 682

Qy        841 TTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 TTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCG 622

Qy        901 GCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        621 GCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGG 562

Qy        961 GGGATACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1001
              ||   |||||||||||||||||||||||||||||||||||
Db        561 GGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 521


RESULT 15
US-10-057-108-1
; Sequence 1, Application US/10057108
; Publication No. US20030061626A1
; GENERAL INFORMATION:
;  APPLICANT: Plaetinck, Geert
;  APPLICANT:  Platteeuw, Christ
;  APPLICANT:  Mortier, Katharine
;  APPLICANT:  Bogaert, Thierry
;  TITLE OF INVENTION: Characterisation of Gene Function using dsRNA
;  TITLE OF INVENTION:  Inhibition
;  FILE REFERENCE: D0590/7003
;  CURRENT APPLICATION NUMBER: US/10/057,108
;  CURRENT FILING DATE:  2002-01-25
;  PRIOR APPLICATION NUMBER: 09/347,311
;  PRIOR FILING DATE: 2001-01-04
;  PRIOR APPLICATION NUMBER: GB 9814536.0
;  PRIOR FILING DATE: 1998-07-03
;  PRIOR APPLICATION NUMBER: GB 9827152.1
;  PRIOR FILING DATE: 1998-12-09
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 3216
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
```

; OTHER INFORMATION: Description of Artificial Sequence:plasmid DNA
US-10-057-108-1

Query Match 35.0%; Score 350.8; DB 9; Length 3216;
Best Local Similarity 88.5%; Pred. No. 7.9e-106;
Matches 355; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

```
Qy        601 TCCAAATGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCCTGTGTGAAA 660
              || | |  |                                        ||||||||||||
Db        785 TCTATAGTGTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAA 844

Qy        661 TTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        845 TTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTG 904

Qy        721 GGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        905 GGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCA 964

Qy        781 GTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        965 GTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGG 1024

Qy        841 TTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1025 TTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCG 1084

Qy        901 GCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1085 GCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGG 1144

Qy        961 GGGATACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1001
              ||   ||||||||||||||||||||||||||||||||||||
Db       1145 GGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA 1185
```

Search completed: July 16, 2003, 18:04:14
Job time : 135.716 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 16, 2003, 03:49:46 ; Search time 2395.68 Seconds
(without alignments)
15998.962 Million cell updates/sec

Title: US-09-981-353-81
Perfect score: 1317
Sequence: 1 aaggaacaaagtaagtccat.....tttttccttaataaaatgtc 1317

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1109.2	84.2	1182	9	AK000224	AK000224 Homo sapi
2	1108	84.1	1176	9	BC029793	BC029793 Homo sapi
c 3	551	41.8	1222	4	SSR236932	AJ236932 Sus scrof
c 4	399	30.3	166793	2	AC090401	AC090401 Homo sapi
5	399	30.3	167934	9	AP001034	AP001034 Homo sapi
6	303.6	23.1	351	6	AX245564	AX245564 Sequence
7	292.6	22.2	342	6	AX245708	AX245708 Sequence
8	291.2	22.1	373	6	AX071292	AX071292 Sequence
9	288.6	21.9	383	6	AX245551	AX245551 Sequence
10	276	21.0	370	6	AX071281	AX071281 Sequence
11	81.2	6.2	852	9	AF350504	AF350504 Homo sapi
12	81.2	6.2	1305	9	BC022895	BC022895 Homo sapi
13	81.2	6.2	1336	6	AX368966	AX368966 Sequence
14	81.2	6.2	1369	9	AF237905	AF237905 Homo sapi
15	79.6	6.0	861	6	AX369167	AX369167 Sequence
16	79.6	6.0	1155	6	AX369165	AX369165 Sequence
17	79.6	6.0	1353	6	AX369163	AX369163 Sequence
18	75.6	5.7	950	10	AF321128	AF321128 Mus muscu
19	75.4	5.7	518	10	AF280401	AF280401 Mus muscu
20	74.8	5.7	862	9	AB013104	AB013104 Homo sapi
21	74.8	5.7	930	9	AF286866	AF286866 Homo sapi
22	74.8	5.7	932	6	AX357469	AX357469 Sequence
23	74.8	5.7	981	9	AF354930	AF354930 Homo sapi
24	74.8	5.7	985	9	AF350502	AF350502 Homo sapi
25	74.8	5.7	1036	6	AX357467	AX357467 Sequence
26	74.8	5.7	1131	9	AF212240	AF212240 Homo sapi
27	74.8	5.7	1152	9	AF237908	AF237908 Homo sapi
28	74.8	5.7	1258	9	AF350503	AF350503 Homo sapi
29	74.8	5.7	1289	9	AF142409	AF142409 Homo sapi
30	74.8	5.7	1433	9	BC022854	BC022854 Homo sapi
31	74.8	5.7	1488	6	AX017505	AX017505 Sequence
32	74.8	5.7	1679	6	AX017610	AX017610 Sequence
33	71.2	5.4	2542	10	RATFCIGA	M22923 Rat IgE Fc
34	71.2	5.4	2545	6	AR123798	AR123798 Sequence
35	71.2	5.4	2545	6	I09687	I09687 Sequence 5
36	70.6	5.4	1025	10	AF237906	AF237906 Mus muscu
37	70.6	5.4	1557	9	BC031610	BC031610 Homo sapi
38	69.2	5.3	708	6	AX101306	AX101306 Sequence
39	69.2	5.3	1010	10	MUSFCERB	J05019 Mouse mast
c 40	68.6	5.2	107211	2	AC111330	AC111330 Rattus no
41	67.4	5.1	1066	10	AB026044	AB026044 Mus muscu
42	61.6	4.7	2292	9	AB071108	AB071108 Macaca fa
43	60.4	4.6	1018	4	OAR318333	AJ318333 Ovis arie
c 44	60	4.6	489	6	AX341797	AX341797 Sequence
45	60	4.6	894	9	HUMIGERB	D10583 Homo sapien

ALIGNMENTS

RESULT 1
AK000224
LOCUS AK000224 1182 bp mRNA linear PRI 19-MAY-2001
DEFINITION Homo sapiens cDNA FLJ20217 fis, clone COLF3334.
ACCESSION AK000224
VERSION AK000224.1 GI:7020167
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens colon mucosa cDNA to mRNA, clone_lib:ColF clone:COLF3334.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Watanabe,K., Kumagai,A., Itakura,S., Yamazaki,M., Tashiro,H., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,

```
            Nakamura,Y., Isogai,T. and Sugano,S.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1182)
  AUTHORS   Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
            Shibahara,T., Tanaka,T. and Nakamura,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science,
            University of Tokyo, Deptment of Virology; Shirokane-dai, 4-6-1,
            Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna1@ims.u-tokyo.ac.jp,
            Tel:81-3-5449-5286, Fax:81-3-5449-5416)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing: Departent of
            Virology and Human Genome Center, Institute of Medical Science,
            University of Tokyo (partly supported by Science and Technology
            Agency).
FEATURES             Location/Qualifiers
     source          1. .1182
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="COLF3334"
                     /tissue_type="colon mucosa"
                     /clone_lib="ColF"
                     /note="cloning vector pME18SFL3"
     gene            1. .1182
                     /gene="MS4A12"
     CDS             58. .861
                     /gene="MS4A12"
                     /codon_start=1
                     /product="MS4A12 protein"
                     /protein_id="BAA91020.1"
                     /db_xref="GI:7020168"
                     /translation="MMSSKPTSHAEVNETIPNPYPPGSFMAPGFQQPLGSINLENQAQ
                     GAQRAQPYGITSPGIFASSQPGQGNIQMINPSVGTAVMNFKEEAKALGVIQIMVGLMH
                     IGFGIVLCLISFSFREVLGFASTAVIGGYPFWGGLSFIISGSLSVSASKELSRCLVKG
                     SLGMNIVSSILAFIGVILLLVDMCINGVAGQDYWAVLSGKGISATLMIFSLLEFFVAC
                     ATAHFANQANTTTNMSVLVIPNMYESNPVTPASSSAPPRCNNYSANAPK"
BASE COUNT      344 a    252 c    245 g    341 t
ORIGIN

  Query Match           84.2%;  Score 1109.2;  DB 9;  Length 1182;
  Best Local Similarity   98.0%;  Pred. No. 9.1e-270;
  Matches 1123;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy        172 TGAACGAAACAAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATC 231
              || |  | | ||  | ||||||  |    |       |   | ||||||||||||||||||||
Db          9 TGGAGCAGAGAAAGAGGAAACATAGAGGTGCCAAAGGAACAAAGACATAATGATGTCATC 68

Qy        232 CAAGCCAACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAG 291
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db         69 CAAGCCAACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAGGCAG 128

Qy        292 CTTTATGGCTCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCA 351
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        129 CTTTATGGCTCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCA 188

Qy        352 GGGTGCTCAGCGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCA 411
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        189 GGGTGCTCAGCGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCA 248

Qy        412 ACCGGGTCAAGGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTT 471
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        249 ACCGGGTCAAGGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTT 308

Qy        472 TAAAGAAGAAGCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGG 531
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        309 TAAAGAAGAAGCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGG 368

Qy        532 TTTTGGAATTGTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTC 591
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        369 TTTTGGAATTGTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTC 428

Qy        592 TACTGCTGTTATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTC 651
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        429 TACTGCTGTTATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTC 488

Qy        652 TCTCTCTGTGTCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAAT 711
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        489 TCTCTCTGTGTCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAAT 548

Qy        712 GAACATTGTTAGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTG 771
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        549 GAACATTGTTAGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTG 608

Qy        772 CATCAATGGGGTAGCTGGCCAAGACTACTGGGCCGTGCTTTCTGGAAAAGGCATTTCAGC 831
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        609 CATCAATGGGGTAGCTGGCCAAGACTACTGGGCCGTGCTTTCTGGAAAAGGCATTTCAGC 668

Qy        832 CACGCTGATGATCTTCTCCCTCTTGGAGTTCTTCGTAGCTTGTGCCACAGCCCATTTTGC 891
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        669 CACGCTGATGATCTTCTCCCTCTTGGAGTTCTTCGTAGCTTGTGCCACAGCCCATTTTGC 728

Qy        892 CAACCAAGCAAACACCACAACCAATATGTCTGTCCTGGTTATTCCAAATATGTATGAAAG 951
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        729 CAACCAAGCAAACACCACAACCAATATGTCTGTCCTGGTTATTCCAAATATGTATGAAAG 788

Qy        952 CAACCCTGTGACACCAGCGTCTTCTTCAGCTCCTCCCAGATGCAACAACTACTCAGCTAA 1011
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        789 CAACCCTGTGACACCAGCGTCTTCTTCAGCTCCTCCCAGATGCAACAACTACTCAGCTAA 848

Qy       1012 TGCCCCTAAATAGTAAAAGAAAAAGGGGTATCAGTCTAATCTCATGGAGAAAAACTACTT 1071
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        849 TGCCCCTAAATAGTAAAAGAAAAAGGGGTATCAGTCTAATCTCATGGAGAAAAACTACTT 908

Qy       1072 GCAAAAACTTCTTAAGAAGATGTCTTTTATTGTCTACAATGATTTCTAGTCTTTAAAAAC 1131
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        909 GCAAAAACTTCTTAAGAAGATGTCTTTTATTGTCTACAATGATTTCTAGTCTTTAAAAAC 968

Qy       1132 TGTGTTTGAGATTTGTTTTTAGGTTGGTCGCTAATGATGGCTGTATCTCCCTTCACTGTC 1191
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        969 TGTGTTTGAGATTTGTTTTTAGGTTGGTCGCTAATGATGGCTGTATCTCCCTTCACTGTC 1028

Qy       1192 TCTTCCTACATTACCACTACTACATGCTGGCAAAGGTGAAGGATCAGAGGACTGAAAAAT 1251
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1029 TCTTCCTACATTACCACTACTACATGCTGGCAAAGGTGAAGGATCAGAGGACTGAAAAAT 1088

Qy       1252 GATTCTGCAACTCTCTTAAAGTTAGAAATGTTTCTGTTCATATTACTTTTTCCTTAATAA 1311
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1089 GATTCTGCAACTCTCTTAAAGTTAGAAATGTTTCTGTTCATATTACTTTTTCCTTAATAA 1148

Qy       1312 AATGTC 1317
              ||||||
Db       1149 AATGTC 1154


RESULT 2
BC029793
LOCUS       BC029793               1176 bp    mRNA    linear   PRI 20-MAY-2002
DEFINITION  Homo sapiens, membrane-spanning 4-domains, subfamily A, member 12
            4-domains, subfamily A, member 7, clone MGC:34288 IMAGE:5185424,
            mRNA, complete cds.
ACCESSION   BC029793
VERSION     BC029793.1  GI:20988642
KEYWORDS    MGC.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1176)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
```

```
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amg@bcm.tmc.edu
            Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
            Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
            Richards, S., Gibbs, R.A.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 50 Row: f Column: 23
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 8923205.
FEATURES              Location/Qualifiers
     source           1. .1176
                      /organism="Homo sapiens"
                      /db_xref="LocusID:54860"
                      /db_xref="taxon:9606"
                      /clone="MGC:34288 IMAGE:5185424"
                      /tissue_type="Colon, Kidney, Stomach, adult, whole pooled"
                      /clone_lib="NIH_MGC_116"
                      /lab_host="DH10B"
                      /note="Vector: pCMV-SPORT6"
     CDS              46. .849
                      /codon_start=1
                      /product="membrane-spanning 4-domains, subfamily A, member
                      12 4-domains, subfamily A, member 7"
                      /protein_id="AAH29793.1"
                      /db_xref="GI:20988643"
                      /translation="MMSSKPTSHAEVNETIPNPYPPSSFMAPGFQQPLGSINLENQAQ
                      GAQRAQPYGITSPGIFASSQPGQGNIQMINPSVGTAVMNFKEEAKALGVIQIMVGLMH
                      IGFGIVLCLISFSFREVLGFASTAVIGGYPFWGGLSFIISGSLSVSASKELSRCLVKG
                      SLGMNIVSSILAFIGVILLLVDMCINGVAGQDYWAVLSGKGISATLMIFSLLEFFVAC
                      ATAHFANQANTTTNMSVLVIPNMYESNPVTPASSSAPPRCNNYSANAPK"
BASE COUNT      347 a    249 c    240 g    340 t
ORIGIN

  Query Match           84.1%;  Score 1108;  DB 9;  Length 1176;
  Best Local Similarity  98.2%;  Pred. No. 1.8e-269;
  Matches 1120;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy        178 AAACAAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATCCAAGCC 237
              | | ||  | |||||| |   |      |   | ||||||||||||||||||||||||||
Db          3 AGAGAAAGAGGAAACATAGAGGTGCCAAAGGAACAAAGACATAATGATGTCATCCAAGCC 62

Qy        238 AACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAGCTTTAT 297
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         63 AACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAGCTTTAT 122

Qy        298 GGCTCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCAGGGTGC 357
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 GGCTCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCAGGGTGC 182

Qy        358 TCAGCGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCAACCGGG 417
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 TCAGCGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCAACCGGG 242

Qy        418 TCAAGGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTTTAAAGA 477
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 TCAAGGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTTTAAAGA 302

Qy        478 AGAAGCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGGTTTTGG 537
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 AGAAGCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGGTTTTGG 362

Qy        538 AATTGTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCTACTGC 597
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 AATTGTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCTACTGC 422

Qy        598 TGTTATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCTCTCTC 657
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 TGTTATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCTCTCTC 482

Qy        658 TGTGTCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATGAACAT 717
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        483 TGTGTCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATGAACAT 542

Qy        718 TGTTAGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGCATCAA 777
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        543 TGTTAGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGCATCAA 602

Qy        778 TGGGGTAGCTGGCCAAGACTACTGGGCCGTGCTTTCTGGAAAAGGCATTTCAGCCACGCT 837
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        603 TGGGGTAGCTGGCCAAGACTACTGGGCCGTGCTTTCTGGAAAAGGCATTTCAGCCACGCT 662

Qy        838 GATGATCTTCTCCCTCTTGGAGTTCTTCGTAGCTTGTGCCACAGCCCATTTTGCCAACCA 897
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        663 GATGATCTTCTCCCTCTTGGAGTTCTTCGTAGCTTGTGCCACAGCCCATTTTGCCAACCA 722

Qy        898 AGCAAACACCACAACCAATATGTCTGTCCTGGTTATTCCAAATATGTATGAAAGCAACCC 957
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        723 AGCAAACACCACAACTAATATGTCTGTCCTGGTTATTCCAAATATGTATGAAAGCAACCC 782

Qy        958 TGTGACACCAGCGTCTTCTTCAGCTCCTCCCAGATGCAACAACTACTCAGCTAATGCCCC 1017
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        783 TGTGACACCAGCGTCTTCTTCAGCTCCTCCCAGATGCAACAACTACTCAGCTAATGCCCC 842

Qy       1018 TAAATAGTAAAAGAAAAAGGGGTATCAGTCTAATCTCATGGAGAAAAACTACTTGCAAAA 1077
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        843 TAAATAGTAAAAGAAAAAGGGGTATCAGTCTAATCTCATGGAGAAAAACTACTTGCAAAA 902

Qy       1078 ACTTCTTAAGAAGATGTCTTTTATTGTCTACAATGATTTCTAGTCTTTAAAAACTGTGTT 1137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         903 ACTTCTTAAGAAGATGTCTTTTATTGTCTACAATGATTTCTAGTCTTTAAAAACTGTGTT 962

Qy       1138 TGAGATTTGTTTTTAGGTTGGTCGCTAATGATGGCTGTATCTCCCTTCACTGTCTCTTCC 1197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         963 TGAGATTTGTTTTTAGGTTGGTCGCTAATGATGGCTGTATCTCCCTTCACTGTCTCTTCC 1022

Qy       1198 TACATTACCACTACTACATGCTGGCAAAGGTGAAGGATCAGAGGACTGAAAAATGATTCT 1257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1023 TACATTACCACTACTACATGCTGGCAAAGGTGAAGGATCAGAGGACTGAAAAATGATTCT 1082

Qy       1258 GCAACTCTCTTAAAGTTAGAAATGTTTCTGTTCATATTACTTTTTCCTTAATAAAATGTC 1317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1083 GCAACTCTCTTAAAGTTAGAAATGTTTCTGTTCATATTACTTTTTCCTTAATAAAATGTC 1142


RESULT 3
SSR236932/c
LOCUS       SSR236932              1222 bp    mRNA    linear   MAM 18-AUG-2000
DEFINITION  Sus scrofa mRNA for hypothetical protein (5'; clone 4B8).
ACCESSION   AJ236932
VERSION     AJ236932.1  GI:4186144
KEYWORDS    .
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 1222)
  AUTHORS   Cirera,S., Wintero,A.K. and Fredholm,M.
  TITLE     Why do we still find anonymous ESTs?
  JOURNAL   Mamm. Genome 11 (8), 689-693 (2000)
  MEDLINE   20380830
   PUBMED   10920241
REFERENCE   2  (bases 1 to 1222)
  AUTHORS   Cirera,S.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (12-JAN-1999) Cirera S., The Royal Veterinary and
            Agricultural University, Department of Animal Science and Animal
            Health, Division of Animal Genetics, Groennegaardsvej 3, 1870
            Frederiksberg C, DENMARK
FEATURES             Location/Qualifiers
     source          1. .1222
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone="4B8"
                     /tissue_type="small intestine"
                     /clone_lib="directionally cloned cDNA in XL1-blue MRF'"
     mRNA            1. .1222
                     /product="hypothetical protein"
BASE COUNT      324 a    263 c    288 g    345 t      2 others
ORIGIN

  Query Match             41.8%;  Score 551;  DB 4;  Length 1222;
  Best Local Similarity   72.1%;  Pred. No. 1.5e-128;
  Matches  836;  Conservative    0;  Mismatches  295;  Indels   28;  Gaps    8;

Qy        173 GAACGAAACAAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATCC 232
              ||  || | |   | ||| | ||  ||      |   | | ||||| ||||||||||||
Db       1175 GAGAGACAGAGAGAGGAAGCCAAGAAGTGCCAAAGGAACAGAGACAAAATGATGTCATCC 1116

Qy        233 AAGCCAACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAGC 292
              |||||||||| | |  ||| ||| ||| ||||| ||  | | ||||||||||| | ||
Db       1115 AAGCCAACAACCTACCCTGGAGTGTATGGAACCACACCTGATCTTTACCCACCAAGTAAC 1056

Qy        293 TTTATGGCTCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCAG 352
              |||||||  ||||| | |||||| || |  || |  ||||||  |  || |||||  ||||
Db       1055 TTTATGGTGCCTGGGTCTCAACAACCCCCAGGGTTCATCAACCCACGAATCCAAGTTCAG 996

Qy        353 GGTGCTCAGCGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCAA 412
               ||   |||    ||           ||||   ||||| ||||||     || ||||||
Db       995 AGTAGCCAGGCACCT---------TTCATCGTCTCTCCAGGAATCCCCAATAACAGTCAG 945

Qy        413 CCGGGTCAAGGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTTT 472
              | || ||||||||||||||||||| ||||||||  ||   |||| |||||   |  ||||||
Db       944 CAGGTTCAAGGAAATATACAAATGGTAAATCCAGGTACAGGAAAAGCAGCCACAAACTTT 885

Qy        473 AAAGAAGAAGCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGGT 532
              |||||||||||||||| |||||||||  |||||||||  || ||||||  ||||||||||
Db       884 AAAGAAGAAGCAAAGACACTAGGGGCAATCCAGATCCTGATTGGATCAATGCACATTGGC 825

Qy        533 TTTGGAATTGTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCT 592
              || |||||  || ||  | |||||    |  | ||| ||    |||| || ||||||||  ||
Db       824 TTCGGAATTATTCTGGGCTTAATGGGCCGCACTTATATGCAAGTTTTGGGTTTTGCTTCC 765

Qy        593 ACTGCTGTTATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCT 652
                 ||| || || | |||||  ||||||||||| |||||  |||||  |||||   ||||   |
Db       764 TTGGCTTTTGTTAGCGGATATCCATTCTGGGGCGGCCTCTCTTTCATTATTACTGGAATT 705

Qy        653 CTCTCTGTGTCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATG 712
              |||| | | |  |||||||||  ||  |||||  |  ||||  | |||  ||||||||||||||
Db       704 CTCTGTATTTTGGCATCCAAGAAGTCTTCCCCTGCTCTGATAAAAAGCAGCCTGGGAATG 645

Qy        713 AACATTGTTAGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGC 772
              |  ||||||||||  ||||  |||||||||||  |||| |||  ||||||||||    ||
Db       644 AGTATTGTTAGTTCCTTCTTTGCCTTCATTGGGATGATCCTGTTGCTGGTGGATGAAAGC 585

Qy        773 ATCAATGGGGTAGCTGGCCAAGACTACTGGGCCGTGCTTTCTGGAAAAGGCATTTCAGCC 832
              |||||  |||  |  | |  ||||||||||||||| |||||||||||||||||||||||||||
Db       584 ATCAACGGGCTGCCCGAGCAAGACTACTGGGCGGTGCTTTCTGGAAAAGGCATTTCAGCC 525

Qy        833 ACGCTGATGATCTTCTCCCTCTTGGAGTTCTTCGTAGCTTGTGCCACAGCCCATTTTGCC 892
              | ||||||  |||||||| |||||||||||||  | || ||||||  |||||||  ||||||||
Db       524 ATGCTGATTATCTTCTCTCTCTTGGAGTTCTGCATAACTTGTGTCACAGCCTATTTTGCC 465

Qy        893 AACCAAGCAAACACCACAACCAATATGTCTGTCCTGGTTATTCCAAATATGTATGAAAGC 952
              | |||  | | |||||  ||| |  ||||||||||| ||||||  | ||||||  || | |
Db       464 AGCCACACCATCACCAACACC-AGGGGTCTGTCCTGGTCATTCCACCTGTGTATGCAAAC 406

Qy        953 AACCCTGTGACACCAGCGTCTTCTTCAGCTCCTCCCAGA---------TGCAACAACTAC 1003
```

```
              |  ||  ||||  |||  |  || |||| |||||||  |             |   |
Db        405 AGTCCCTTGACGACAGGATTGTCCTCAGTTCCTCCCCCAGTTCCGGCCACTTAGCTCACA 346

Qy       1004 TCAGCTAATGCCCCTAAATAGTAAAAGAAAAAGGGGTATCAGTCTAATCTCATGGAGAAA 1063
               || ||||| ||  ||  || |||||||| |||  | |||   | | |     || | |
Db        345 CCACCTAATACCAGTAGGTATTAAAAGAAGAAGAAGCATCCTGCAATTTCATGGGGGTAG 286

Qy       1064 AACTACT-TGCAAAAACTTCTTAAGAAGATGTCTTTTATTGTCTACAATGATTTCTAGTC 1122
              |||  || |||| |||||||||||||| ||   ||    ||  || |||||  ||| ||
Db        285 AACCGCTCTGCAGAAACTTCTTAAGAAGTTGGTCTTCGCCGTTCACGATGATTGCTAATC 226

Qy       1123 TTTAAAAACTGTGTTTGAGATTTGTTTTTAGGTTGGTCGCTAATGATGGCTGTATCTCCC 1182
              |||||| |||||||| || ||   || | |||| |||   |||||   || |||   ||
Db        225 TTTAAAGACTGTGTTTCAGGTTCTGTTCTTGGTTTGTCATGAATGAAAACTCTATGGACC 166

Qy       1183 T-----TCACTGTCTCTTCCTACATTACCACTACTACATGCTGGCAAAGGTGAAGGATCA 1237
              |       || ||| |||||||||| | ||||||| || ||||||||||||||||||
Db        165 TCTCTACCAGTGTTTCTTCCTACA-TGCCACTACCACGTGCTGGCAAAGGTGAAGGGCTG 107

Qy       1238 GAGGACTGAAAAATGATTCTGCAACTCTCTTAAAGTTAGAAATGTTTCTGTTCATATTAC 1297
              ||||||||  |||||| ||||||||||||||||||||||||||| || ||||| |  ||||
Db        106 GAGGACTG-GAAATGACTCTGCAACTCTCTTAAAGTTAGAAAT-TTCCTGTTAAAGTTAC 49

Qy       1298 TTTTTCCTTAATAAAATGT 1316
                 || ||| ||||||||||
Db         48 CCTTCCCTCAATAAAATGT 30


RESULT 4
AC090401/c
LOCUS       AC090401              166793 bp    DNA     linear   HTG 10-APR-2002
DEFINITION  Homo sapiens chromosome 11 clone RP11-729B4 map 11, *** SEQUENCING
            IN PROGRESS ***, 1 ordered piece.
ACCESSION   AC090401
VERSION     AC090401.5  GI:20128179
KEYWORDS    HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 166793)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-729B4
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 166793)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 166793)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
```

Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (10-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT On Apr 10, 2002 this sequence version replaced gi:19683210.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-------------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-------------- Project Information
Center project name: L12709
Center clone name: 729_B_4

* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 166793: contig of 166793 bp in length.

FEATURES Location/Qualifiers
source 1. .166793
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone="RP11-729B4"
/clone_lib="RPCI-11 Human Male BAC"

BASE COUNT 46459 a 31834 c 32914 g 55586 t
ORIGIN

Query Match 30.3%; Score 399; DB 2; Length 166793;
Best Local Similarity 100.0%; Pred. No. 4.3e-90;
Matches 399; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       919 GTCTGTCCTGGTTATTCCAAATATGTATGAAAGCAACCCTGTGACACCAGCGTCTTCTTC 978
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     66730 GTCTGTCCTGGTTATTCCAAATATGTATGAAAGCAACCCTGTGACACCAGCGTCTTCTTC 66671

Qy       979 AGCTCCTCCCAGATGCAACAACTACTCAGCTAATGCCCCTAAATAGTAAAAGAAAAAGGG 1038
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     66670 AGCTCCTCCCAGATGCAACAACTACTCAGCTAATGCCCCTAAATAGTAAAAGAAAAAGGG 66611

Qy      1039 GTATCAGTCTAATCTCATGGAGAAAAACTACTTGCAAAAACTTCTTAAGAAGATGTCTTT 1098
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     66610 GTATCAGTCTAATCTCATGGAGAAAAACTACTTGCAAAAACTTCTTAAGAAGATGTCTTT 66551

Qy      1099 TATTGTCTACAATGATTTCTAGTCTTTAAAAACTGTGTTTGAGATTTGTTTTTAGGTTGG 1158
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     66550 TATTGTCTACAATGATTTCTAGTCTTTAAAAACTGTGTTTGAGATTTGTTTTTAGGTTGG 66491

Qy      1159 TCGCTAATGATGGCTGTATCTCCCTTCACTGTCTCTTCCTACATTACCACTACTACATGC 1218
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     66490 TCGCTAATGATGGCTGTATCTCCCTTCACTGTCTCTTCCTACATTACCACTACTACATGC 66431

Qy      1219 TGGCAAAGGTGAAGGATCAGAGGACTGAAAAATGATTCTGCAACTCTCTTAAAGTTAGAA 1278
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     66430 TGGCAAAGGTGAAGGATCAGAGGACTGAAAAATGATTCTGCAACTCTCTTAAAGTTAGAA 66371

Qy      1279 ATGTTTCTGTTCATATTACTTTTTCCTTAATAAAATGTC 1317
            |||||||||||||||||||||||||||||||||||||||
Db     66370 ATGTTTCTGTTCATATTACTTTTTCCTTAATAAAATGTC 66332
```

RESULT 5
AP001034
LOCUS AP001034 167934 bp DNA linear PRI 06-NOV-2001
DEFINITION Homo sapiens genomic DNA, chromosome 11q clone:RP11-729B4, complete sequence.
ACCESSION AP001034
VERSION AP001034.5 GI:16751479
KEYWORDS HTG.
SOURCE Homo sapiens DNA, clone:RP11-729B4.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Homo sapiens genomic DNA
JOURNAL Published Only in Database (2000)

REFERENCE 2 (bases 1 to 167934)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (06-JAN-2000) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)

COMMENT On Nov 5, 2001 this sequence version replaced gi:11176992.

FEATURES Location/Qualifiers
source 1. .167934
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-729B4"

BASE COUNT 56061 a 33108 c 32009 g 46756 t
ORIGIN

Query Match 30.3%; Score 399; DB 9; Length 167934;
Best Local Similarity 100.0%; Pred. No. 4.3e-90;
Matches 399; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       919 GTCTGTCCTGGTTATTCCAAATATGTATGAAAGCAACCCTGTGACACCAGCGTCTTCTTC 978
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    100060 GTCTGTCCTGGTTATTCCAAATATGTATGAAAGCAACCCTGTGACACCAGCGTCTTCTTC 100119

Qy       979 AGCTCCTCCCAGATGCAACAACTACTCAGCTAATGCCCCTAAATAGTAAAAGAAAAAGGG 1038
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    100120 AGCTCCTCCCAGATGCAACAACTACTCAGCTAATGCCCCTAAATAGTAAAAGAAAAAGGG 100179

Qy      1039 GTATCAGTCTAATCTCATGGAGAAAAACTACTTGCAAAAACTTCTTAAGAAGATGTCTTT 1098
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    100180 GTATCAGTCTAATCTCATGGAGAAAAACTACTTGCAAAAACTTCTTAAGAAGATGTCTTT 100239

Qy      1099 TATTGTCTACAATGATTTCTAGTCTTTAAAAACTGTGTTTGAGATTTGTTTTTAGGTTGG 1158
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    100240 TATTGTCTACAATGATTTCTAGTCTTTAAAAACTGTGTTTGAGATTTGTTTTTAGGTTGG 100299
```

```
Qy       1159 TCGCTAATGATGGCTGTATCTCCCTTCACTGTCTCTTCCTACATTACCACTACTACATGC 1218
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     100300 TCGCTAATGATGGCTGTATCTCCCTTCACTGTCTCTTCCTACATTACCACTACTACATGC 100359

Qy       1219 TGGCAAAGGTGAAGGATCAGAGGACTGAAAAATGATTCTGCAACTCTCTTAAAGTTAGAA 1278
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     100360 TGGCAAAGGTGAAGGATCAGAGGACTGAAAAATGATTCTGCAACTCTCTTAAAGTTAGAA 100419

Qy       1279 ATGTTTCTGTTCATATTACTTTTTCCTTAATAAAATGTC 1317
              |||||||||||||||||||||||||||||||||||||||
Db     100420 ATGTTTCTGTTCATATTACTTTTTCCTTAATAAAATGTC 100458


RESULT 6
AX245564
LOCUS       AX245564                351 bp    DNA     linear   PAT 28-SEP-2001
DEFINITION  Sequence 494 from Patent WO0166753.
ACCESSION   AX245564
VERSION     AX245564.1  GI:15860238
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 351)
  AUTHORS   Williams,L.T., Escobedo,J., Innis,M.A., Garcia,P.D.,
            Sudduth-Klinger,J., Reinhard,C., Randazzo,F., Kennedy,G.C., Pot,D.,
            Kassam,A., Lamson,G., Drmanac,R., Crkvenjakov,R., Dickson,M.,
            Drmanac,S., Labat,I., Leshkowitz,D., Kita,D., Garcia,V. and
            Stache-Crain,B.
  TITLE     Human genes and gene expression products
  JOURNAL   Patent: WO 0166753-A 494 13-SEP-2001;
            Chiron Corporation (US) ; Hyseq Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .351
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      123 a     79 c     81 g     68 t
ORIGIN

  Query Match           23.1%;  Score 303.6;  DB 6;  Length 351;
  Best Local Similarity   93.0%;  Pred. No. 5.7e-66;
  Matches  318;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy        172 TGAACGAAACAAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATC 231
              || |  | | ||  | |||||| |   |       |   | ||||||||||||||||||||
Db         10 TGGAGCAGAGAAAGAGGAAACATAGAGGTGCCAAAGGAACAAAGACATAATGATGTCATC 69

Qy        232 CAAGCCAACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAG 291
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         70 CAAGCCAACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAG 129

Qy        292 CTTTATGGCTCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCA 351
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        130 CTTTATGGCTCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCA 189

Qy        352 GGGTGCTCAGCGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCA 411
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        190 GGGTGCTCAGCGTGCTCAGCCCTATGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCA 249

Qy        412 ACCGGGTCAAGGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTT 471
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        250 ACCGGGTCAAGGAAATATATAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTT 309

Qy        472 TAAAGAAGAAGCAAAGGCACTAGGGGTGATCCAGATCATGGT 513
              |||||||||||||||||||||||||||||||||||||||||||
Db        310 TAAAGAAGAAGCAAAGGCACTAGGGGTGATCCAGATCATGGT 351


RESULT 7
AX245708
LOCUS       AX245708                342 bp    DNA     linear   PAT 28-SEP-2001
DEFINITION  Sequence 638 from Patent WO0166753.
ACCESSION   AX245708
VERSION     AX245708.1  GI:15860382
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 342)
  AUTHORS   Williams,L.T., Escobedo,J., Innis,M.A., Garcia,P.D.,
            Sudduth-Klinger,J., Reinhard,C., Randazzo,F., Kennedy,G.C., Pot,D.,
            Kassam,A., Lamson,G., Drmanac,R., Crkvenjakov,R., Dickson,M.,
            Drmanac,S., Labat,I., Leshkowitz,D., Kita,D., Garcia,V. and
            Stache-Crain,B.
  TITLE     Human genes and gene expression products
  JOURNAL   Patent: WO 0166753-A 638 13-SEP-2001;
            Chiron Corporation (US) ; Hyseq Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .342
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      119 a     78 c     78 g     65 t      2 others
ORIGIN

  Query Match           22.2%;  Score 292.6;  DB 6;  Length 342;
  Best Local Similarity   92.5%;  Pred. No. 3.4e-63;
  Matches  307;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy        172 TGAACGAAACAAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATC 231
              || |  | | ||  | |||||| |   |       |   | ||||||||||||||||||||
Db         10 TGGAGCAGAGAAAGAGGAAACATAGAGGTGCCAAAGGAACAAAGACATAATGATGTCATC 69

Qy        232 CAAGCCAACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAG 291
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         70 CAAGCCAACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAG 129

Qy        292 CTTTATGGCTCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCA 351
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        130 CTTTATGGCTCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCA 189

Qy        352 GGGTGCTCAGCGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCA 411
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        190 GGGTGCTCAGCGTGCTCAGCCCTATGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCA 249

Qy        412 ACCGGGTCAAGGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTT 471
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        250 ACCGGGTCAAGGAAATATATAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTT 309

Qy        472 TAAAGAAGAAGCAAAGGCACTAGGGGTGATCC 503
              |||||||||||| ||||||||||||||||||
Db        310 TAAAGAAGAAGCANAGGCACTAGGGGTGATCC 341


RESULT 8
AX071292
LOCUS       AX071292                373 bp    DNA     linear   PAT 25-JAN-2001
DEFINITION  Sequence 1764 from Patent WO0102568.
ACCESSION   AX071292
VERSION     AX071292.1  GI:12581643
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 373)
  AUTHORS   Williams,L.T., Escobedo,J., Innis,M.A., Garcia,P.D., Klinger,J.,
            Kassam,A., Reinhard,C., Randazzo,F., Kennedy,G.C., Pot,D.,
            Lamson,G., Drmanac,R., Crkenjakov,R., Drmanac,S., Dickson,M.,
            Labat,I., Leshkowitiz,D., Kita,D., Garcia,V. and Strache-Crain,B.
  TITLE     Human genes and gene expression products
  JOURNAL   Patent: WO 0102568-A 1764 11-JAN-2001;
            CHIRON CORPORATION (US) ; HYSEQ, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .373
```

```
                 /organism="Homo sapiens"
                 /db_xref="taxon:9606"
BASE COUNT      127 a     84 c     90 g     72 t
ORIGIN

  Query Match           22.1%;  Score 291.2;  DB 6;  Length 373;
  Best Local Similarity  91.7%;  Pred. No. 7.8e-63;
  Matches  308;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy        172 TGAACGAAACAAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATC 231
              || |  | | ||  | ||| || |   |       |   | | ||||||||||||||||||
Db         37 TGGAGCAGAGAAAGAGGAATCATAGAGGTGCCAAAGGAACACAGACATAATGATGTCATC 96

Qy        232 CAAGCCAACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAG 291
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         97 CAAGCCAACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAG 156

Qy        292 CTTTATGGCTCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCA 351
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        157 CTTTATGGCTCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCA 216

Qy        352 GGGTGCTCAGCGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCA 411
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        217 GGGTGCTCAGCGTGCTCAGCCCTATGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCA 276

Qy        412 ACCGGGTCAAGGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTT 471
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        277 ACCGGGTCAAGGAAATATATAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTT 336

Qy        472 TAAAGAAGAAGCAAAGGCACTAGGGGTGATCCAGAT 507
              ||||||||||| ||||||||||||||||||||||||
Db        337 TAAAGAAGAAGCTTAGGCACTAGGGGTGATCCAGAT 372


RESULT 9
AX245551
LOCUS       AX245551                383 bp    DNA     linear   PAT 28-SEP-2001
DEFINITION  Sequence 481 from Patent WO0166753.
ACCESSION   AX245551
VERSION     AX245551.1  GI:15860225
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 383)
  AUTHORS   Williams,L.T., Escobedo,J., Innis,M.A., Garcia,P.D.,
            Sudduth-Klinger,J., Reinhard,C., Randazzo,F., Kennedy,G.C., Pot,D.,
            Kassam,A., Lamson,G., Drmanac,R., Crkvenjakov,R., Dickson,M.,
            Drmanac,S., Labat,I., Leshkowitz,D., Kita,D., Garcia,V. and
            Stache-Crain,B.
  TITLE     Human genes and gene expression products
  JOURNAL   Patent: WO 0166753-A 481 13-SEP-2001;
            Chiron Corporation (US) ; Hyseq Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .383
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      129 a     82 c     89 g     82 t      1 others
ORIGIN

  Query Match           21.9%;  Score 288.6;  DB 6;  Length 383;
  Best Local Similarity  90.1%;  Pred. No. 3.5e-62;
  Matches  309;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy        172 TGAACGAAACAAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATC 231
              || |  | | ||  ||||| || |   |       |   |  ||||||||||||||| |
Db         41 TGGAGCACAGAAAGAAGAATCATAGAGGTGCCAAAGGAACTTAGACATAATGATGTCGTT 100

Qy        232 CAAGCCAACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAG 291
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        101 CAAGCCAACAAGCCAAGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAG 160

Qy        292 CTTTATGGCTCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCA 351
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db        161 TTTTATGGCTCCTGGATTTCAACAGGCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCA 220

Qy        352 GGGTGCTCAGCGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCA 411
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        221 TGGTGCTCAGCGTGCTCAGCCCTATGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCA 280

Qy        412 ACCGGGTCAAGGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTT 471
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        281 ACCGGGTCAAGGAAATATATAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTT 340

Qy        472 TAAAGAAGAAGCAAAGGCACTAGGGGTGATCCAGATCATGGTT 514
              |||||||||||||||||||| |||||||||| |||||||||
Db        341 TAAAGAAGAAGCAAAGGCACTATGGGTGATCCACATCATGGTT 383


RESULT 10
AX071281
LOCUS       AX071281                370 bp    DNA     linear   PAT 25-JAN-2001
DEFINITION  Sequence 1753 from Patent WO0102568.
ACCESSION   AX071281
VERSION     AX071281.1  GI:12581632
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 370)
  AUTHORS   Williams,L.T., Escobedo,J., Innis,M.A., Garcia,P.D., Klinger,J.,
            Kassam,A., Reinhard,C., Randazzo,F., Kennedy,G.C., Pot,D.,
            Lamson,G., Drmanac,R., Crkenjakov,R., Drmanac,S., Dickson,M.,
            Labat,I., Leshkowitiz,D., Kita,D., Garcia,V. and Strache-Crain,B.
  TITLE     Human genes and gene expression products
  JOURNAL   Patent: WO 0102568-A 1753 11-JAN-2001;
            CHIRON CORPORATION (US) ; HYSEQ, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .370
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      131 a     83 c     89 g     67 t
ORIGIN

  Query Match           21.0%;  Score 276;  DB 6;  Length 370;
  Best Local Similarity  92.1%;  Pred. No. 5.4e-59;
  Matches  291;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy        172 TGAACGAAACAAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATC 231
              || |  | | ||  | |||||| |   |       |   | ||||||||||||||||||
Db         37 TGGAGCAGAGAAAGAGGAAACATAGAGGTGCCAAAGGAACAAAGACATAATGATGTCATC 96

Qy        232 CAAGCCAACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAG 291
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         97 CAAGCCAACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAG 156

Qy        292 CTTTATGGCTCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCA 351
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        157 CTTTATGGCTCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCA 216

Qy        352 GGGTGCTCAGCGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCA 411
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        217 GGGTGCTCAGCGTGCTCAGCCCTATGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCA 276

Qy        412 ACCGGGTCAAGGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTT 471
              |||||| |||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        277 ACCGGGGCAAGGAAATATATAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTT 336

Qy        472 TAAAGAAGAAGCAAAG 487
              ||||||||||||||||
Db        337 TAAAGAAGAAGCAAAG 352


RESULT 11
```

```
AF350504
LOCUS       AF350504                852 bp    mRNA    linear   PRI 01-JAN-2002
DEFINITION  Homo sapiens four-span transmembrane protein 4 (4SPAN4) mRNA,
            complete cds.
ACCESSION   AF350504
VERSION     AF350504.1  GI:18028935
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 852)
  AUTHORS   Hulett,M.D., Pagler,E., Hogarth,M., Eyre,H., Baker,E., Crawford,J.,
            Sutherland,G.R. and Parish,C.R.
  TITLE     Isolation of a family of hematopoietic-expressed four-transmembrane
            genes related to CD20 and FcepsilonRI-beta
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 852)
  AUTHORS   Hulett,M.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-FEB-2001) Immunology and Cell Biology, John Curtin
            School of Medical Research, Mills Road, Canberra, ACT 2601,
            Australia
FEATURES             Location/Qualifiers
     source          1. .852
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q13"
                     /tissue_type="placenta"
     gene            1. .852
                     /gene="4SPAN4"
     CDS             46. .798
                     /gene="4SPAN4"
                     /codon_start=1
                     /product="four-span transmembrane protein 4"
                     /protein_id="AAL56224.1"
                     /db_xref="GI:18028936"
                     /translation="MNSMTSAVPVANSVLVVAPHNGYPVTPGIMSHVPLYPNSQPQVH
                     LVPGNPPSLVSNVNGQPVQKALKEGKTLGAIQIIIGLAHIGLGSIMATVLVGEYLSIS
                     FYGGFPFWGGLWFIISGSLSVAAENQPYSYCLLSGSLGLNIVSAICSAVGVILFITDL
                     SIPHPYAYPDYYPYAWGVNPGMAISGVLLVFCLLEFGIACASSHFGCQLVCCQSSNVS
                     VIYPNIYAANPVITPEPVTSPPSYSSEIQANK"
BASE COUNT      195 a    230 c    204 g    223 t
ORIGIN

  Query Match             6.2%;  Score 81.2;  DB 9;  Length 852;
  Best Local Similarity   58.3%;  Pred. No. 1e-09;
  Matches  190;  Conservative    0;  Mismatches  118;  Indels   18;  Gaps    2;

Qy        590 TCTACTGCTGTTATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGC 649
              |||| | |  |    || || |  || |||||||| ||| | |  ||||| || || ||
Db        343 TCTATTTCATTCTACGGAGGCTTTCCCTTCTGGGGAGGCTTGTGGTTTATCATTTCAGGA 402

Qy        650 TCTCTCTCTGTGTCAGCATCCAAGGAGCT---TTCCCGTTGTCTGGTGAAAGGCAGCCTG 706
              |||||||| ||| |||||   ||  |||    |||   ||| ||| ||   ||||| ||
Db        403 TCTCTCTCCGTGGCAGCAGAAAATCAGCCATATTCTTATTGCCTGCTGTCTGGCAGTTTG 462

Qy        707 GGAATGAACATTGTTAGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGAT 766
              ||  ||||||| ||·||| | ||||   |     ||||||| || ||  |  |    |||
Db        463 GGCTTGAACATCGTCAGTGCAATCTGCTCTGCAGTTGGAGTCATACTCTTCATCACAGAT 522

Qy        767 ATGTGCAT--------------CAATGGGGTAGCTGGCCAAGACTACTGGGCCGTGCTT 811
                | | ||             | |    |    |   |   ||  |||||  |||
Db        523 CTAAGTATTCCCCACCCATATGCCTACCCCGACTATTATCCTTACGCCTGGGGTGTGAAC 582

Qy        812 TCTGGAAAAGGCATTTCAGCCACGCTGATGATCTTCTCCCTCTTGGAGTTCTTCGTAGCT 871
              ||||||  |  ||||| | |  |||| || |||||| |||| |||||||   | | ||
Db        583 CCTGGAATGGCGATTTCTGGCGTGCTGCTGGTCTTCTGCCTCCTGGAGTTTGGCATCGCA 642

Qy        872 TGTGCCACAGCCCATTTTGCCAACCA 897
              || ||  |  |||| |||| |  |||
Db        643 TGCGCATCTTCCCACTTTGGCTGCCA 668
```

```
RESULT 12
BC022895
LOCUS       BC022895                1305 bp    mRNA    linear   PRI 07-FEB-2002
DEFINITION  Homo sapiens, membrane-spanning 4-domains, subfamily A, member 8B,
            clone MGC:27396 IMAGE:4691480, mRNA, complete cds.
ACCESSION   BC022895
VERSION     BC022895.1  GI:18605601
KEYWORDS    MGC.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1305)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-FEB-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
            Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 38 Row: k Column: 6
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 13899294.
FEATURES             Location/Qualifiers
     source          1. .1305
                     /organism="Homo sapiens"
                     /db_xref="LocusID:83661"
                     /db_xref="taxon:9606"
                     /clone="MGC:27396 IMAGE:4691480"
                     /tissue_type="Lung"
                     /clone_lib="NIH_MGC_77"
                     /lab_host="DH10B"
                     /note="Vector: pDNR-LIB"
     CDS             129. .881
                     /codon_start=1
                     /product="membrane-spanning 4-domains, subfamily A, member
                     8B"
                     /protein_id="AAH22895.1"
                     /db_xref="GI:18605602"
                     /translation="MNSMTSAVPVANSVLVVAPHNGYPVTPGIMSHVPLYPNSQPQVH
                     LVPGNPPSLVSNVNGQPVQKALKEGKTLGAIQIIIGLARIGLGSIMATVLVGEYLSIS
                     FYGGFPFWGGLWFIISGSLSVAAENQPYSYCLLSGSLGLNIVSAICSAVGVILFITDL
                     SIPHPYAYPDYYPYAWGVNPGMAISGVLLVFCLLEFGIACASSHFGCQLVCCQSSNVS
                     VIYPNIYAANPVITPEPVTSPPSYSSEIQANK"
BASE COUNT      340 a    341 c    287 g    337 t
ORIGIN

  Query Match             6.2%;  Score 81.2;  DB 9;  Length 1305;
  Best Local Similarity   58.3%;  Pred. No. 1e-09;
  Matches  190;  Conservative    0;  Mismatches  118;  Indels   18;  Gaps    2;

Qy        590 TCTACTGCTGTTATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGC 649
              |||| | |  |    || || |  || |||||||| ||| | |  ||||| || || ||
Db        426 TCTATTTCATTCTACGGAGGCTTTCCCTTCTGGGGAGGCTTGTGGTTTATCATTTCAGGA 485

Qy        650 TCTCTCTCTGTGTCAGCATCCAAGGAGCT---TTCCCGTTGTCTGGTGAAAGGCAGCCTG 706
              |||||||| ||| |||||   ||  |||    |||   ||| ||| ||   ||||| ||
Db        486 TCTCTCTCCGTGGCAGCAGAAAATCAGCCATATTCTTATTGCCTGCTGTCTGGCAGTTTG 545
```

```
Qy        707 GGAATGAACATTGTTAGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGAT 766
              ||  |||||||  ||  |||  |  ||||    |     |||||||  ||  ||   |.  |    |||
Db        546 GGCTTGAACATCGTCAGTGCAATCTGCTCTGCAGTTGGAGTCATACTCTTCATCACAGAT 605

Qy        767 ATGTGCAT---------------CAATGGGGTAGCTGGCCAAGACTACTGGGCCGTGCTT 811
              |  | ||                | |    |    |   |   ||  |||||  |||
Db        606 CTAAGTATTCCCCACCCATATGCCTACCCCGACTATTATCCTTACGCCTGGGGTGTGAAC 665

Qy        812 TCTGGAAAAGGCATTTCAGCCACGCTGATGATCTTCTCCCTCTTGGAGTTCTTCGTAGCT 871
              ||||||  |  |||||  | |  ||||  ||  ||||||  ||||  |||||||   | | ||
Db        666 CCTGGAATGGCGATTTCTGGCGTGCTGCTGGTCTTCTGCCTCCTGGAGTTTGGCATCGCA 725

Qy        872 TGTGCCACAGCCCATTTTGCCAACCA 897
              ||  ||  |  ||||  ||||  |  |||
Db        726 TGCGCATCTTCCCACTTTGGCTGCCA 751


RESULT 13
AX368966
LOCUS       AX368966               1336 bp    DNA     linear   PAT 16-FEB-2002
DEFINITION  Sequence 1676 from Patent WO0204514.
ACCESSION   AX368966
VERSION     AX368966.1  GI:18857032
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Wang,T., Watanabe,Y., Henderson,R.A., Johnson,J.C., Retter,M.W.,
            Marnerakis,M., Carter,D., Fanger,G.R., Vedvick,T.S., Bangur,C.S.,
            Mcnabb,A., Fanger,N., Switzer,A., Mcneill,P.D. and Clapper,J.D.
  TITLE     Compositions and methods for the therapy and diagnosis of lung
            cancer
  JOURNAL   Patent: WO 0204514-A 1676 17-JAN-2002;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .1336
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      336 a    354 c    296 g    350 t
ORIGIN

  Query Match             6.2%;  Score 81.2;  DB 6;  Length 1336;
  Best Local Similarity   58.3%;  Pred. No. 1e-09;
  Matches  190;  Conservative    0;  Mismatches  118;  Indels   18;  Gaps    2;

Qy        590 TCTACTGCTGTTATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGC 649
              ||||  |  |   |    || ||  |  || |||||||| ||| |  |  ||||| || ||  ||
Db        471 TCTATTTCATTCTACGGAGGCTTTCCCTTCTGGGGAGGCTTGTGGTTTATCATTTCAGGA 530

Qy        650 TCTCTCTCTGTGTCAGCATCCAAGGAGCT---TTCCCGTTGTCTGGTGAAAGGCAGCCTG 706
              |||||||| ||| |||||   ||  |||    |||   ||| ||| ||  |||||  ||
Db        531 TCTCTCTCCGTGGCAGCAGAAAATCAGCCATATTCTTATTGCCTGCTGTCTGGCAGTTTG 590

Qy        707 GGAATGAACATTGTTAGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGAT 766
              ||  |||||||  ||  ||| |  ||||    |     |||||||  ||  ||   |.  |    |||
Db        591 GGCTTGAACATCGTCAGTGCAATCTGCTCTGCAGTTGGAGTCATACTCTTCATCACAGAT 650

Qy        767 ATGTGCAT---------------CAATGGGGTAGCTGGCCAAGACTACTGGGCCGTGCTT 811
              |  | ||                | |    |    |   |   ||  |||||  |||
Db        651 CTAAGTATTCCCCACCCATATGCCTACCCCGACTATTATCCTTACGCCTGGGGTGTGAAC 710

Qy        812 TCTGGAAAAGGCATTTCAGCCACGCTGATGATCTTCTCCCTCTTGGAGTTCTTCGTAGCT 871
              ||||||  |  |||||  | |  ||||  ||  ||||||  ||||  |||||||   | | ||
Db        711 CCTGGAATGGCGATTTCTGGCGTGCTGCTGGTCTTCTGCCTCCTGGAGTTTGGCATCGCA 770

Qy        872 TGTGCCACAGCCCATTTTGCCAACCA 897
              ||  ||  |  ||||  ||||  |  |||
Db        771 TGCGCATCTTCCCACTTTGGCTGCCA 796


RESULT 14
AF237905
LOCUS       AF237905               1369 bp    mRNA    linear   PRI 17-APR-2001
DEFINITION  Homo sapiens MS4A8B protein mRNA, complete cds.
ACCESSION   AF237905
VERSION     AF237905.1  GI:13649389
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1369)
  AUTHORS   Liang,Y. and Tedder,T.F.
  TITLE     Identification of a CD20-, FcepsilonRIbeta-, and HTm4-related gene
            family: sixteen new MS4A family members expressed in human and
            mouse
  JOURNAL   Genomics 72 (2), 119-127 (2001)
  MEDLINE   21295030
   PUBMED   11401424
REFERENCE   2  (bases 1 to 1369)
  AUTHORS   Liang,Y. and Tedder,T.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-FEB-2000) Immunology, Duke Medical Center, Research
            Dr., Durham, NC 27710, USA
FEATURES             Location/Qualifiers
     source          1. .1369
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q12-q13"
                     /tissue_type="brain"
                     /dev_stage="infant"
     CDS             204. .956
                     /codon_start=1
                     /product="MS4A8B protein"
                     /protein_id="AAK37414.1"
                     /db_xref="GI:13649390"
                     /translation="MNSMTSAVPVANSVLVVAPHNGYPVTPGIMSHVPLYPNSQPQVH
                     LVPGNPPSLVSNVNGQPVQKALKEGKTLGAIQIIIGLAHIGLGSIMATVLVGEYLSIS
                     FYGGFPFWGGLWFIISGSLSVAAENQPYSYCLLSGSLGLNIVSAICSAVGVILFITDL
                     SIPHPYAYPDYYPYAWGVNPGMAISGVLLVFCLLEFGIACASSHFGCQLVCCQSSNVS
                     VIYPNIYAANPVITPEPVTSPPSYSSEIQANK".
BASE COUNT      360 a    357 c    300 g    352 t
ORIGIN

  Query Match             6.2%;  Score 81.2;  DB 9;  Length 1369;
  Best Local Similarity   58.3%;  Pred. No. 1e-09;
  Matches  190;  Conservative    0;  Mismatches  118;  Indels   18;  Gaps    2;

Qy        590 TCTACTGCTGTTATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGC 649
              ||||  |  |   |    || ||  |  || |||||||| ||| |  |  ||||| || ||  ||
Db        501 TCTATTTCATTCTACGGAGGCTTTCCCTTCTGGGGAGGCTTGTGGTTTATCATTTCAGGA 560

Qy        650 TCTCTCTCTGTGTCAGCATCCAAGGAGCT---TTCCCGTTGTCTGGTGAAAGGCAGCCTG 706
              |||||||| ||| |||||   ||  |||    |||   ||| ||| ||  |||||  ||
Db        561 TCTCTCTCCGTGGCAGCAGAAAATCAGCCATATTCTTATTGCCTGCTGTCTGGCAGTTTG 620

Qy        707 GGAATGAACATTGTTAGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGAT 766
              ||  |||||||  ||  ||| |  ||||    |     |||||||  ||  ||   |.  |    |||
Db        621 GGCTTGAACATCGTCAGTGCAATCTGCTCTGCAGTTGGAGTCATACTCTTCATCACAGAT 680

Qy        767 ATGTGCAT---------------CAATGGGGTAGCTGGCCAAGACTACTGGGCCGTGCTT 811
              |  | ||                | |    |    |   |   ||  |||||  |||
Db        681 CTAAGTATTCCCCACCCATATGCCTACCCCGACTATTATCCTTACGCCTGGGGTGTGAAC 740

Qy        812 TCTGGAAAAGGCATTTCAGCCACGCTGATGATCTTCTCCCTCTTGGAGTTCTTCGTAGCT 871
              ||||||  |  |||||  | |  ||||  ||  ||||||  ||||  |||||||   | | ||
Db        741 CCTGGAATGGCGATTTCTGGCGTGCTGCTGGTCTTCTGCCTCCTGGAGTTTGGCATCGCA 800

Qy        872 TGTGCCACAGCCCATTTTGCCAACCA 897
              ||  ||  |  ||||  ||||  |  |||
Db        801 TGCGCATCTTCCCACTTTGGCTGCCA 826
```

```
RESULT 15
AX369167
LOCUS       AX369167                861 bp    DNA     linear   PAT 16-FEB-2002
DEFINITION  Sequence 1877 from Patent WO0204514.
ACCESSION   AX369167
VERSION     AX369167.1  GI:18857177
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Wang,T., Watanabe,Y., Henderson,R.A., Johnson,J.C., Retter,M.W.,
            Marnerakis,M., Carter,D., Fanger,G.R., Vedvick,T.S., Bangur,C.S.,
            Mcnabb,A., Fanger,N., Switzer,A., Mcneill,P.D. and Clapper,J.D.
  TITLE     Compositions and methods for the therapy and diagnosis of lung
            cancer
  JOURNAL   Patent: WO 0204514-A 1877 17-JAN-2002;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .861
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      186 a    244 c    209 g    222 t
ORIGIN

  Query Match             6.0%;  Score 79.6;  DB 6;  Length 861;
  Best Local Similarity   58.0%;  Pred. No. 2.6e-09;
  Matches  189;  Conservative    0;  Mismatches  119;  Indels   18;  Gaps    2;

Qy        590 TCTACTGCTGTTATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGC 649
              |||| | |  |    || || |  || ||||||| ||| | |  ||||| || || |
Db        406 TCTATTTCATTCTACGGAGGCTTTCCCTTCTGGGGAGGCTTGTGGTTTATCATTTCAGAA 465

Qy        650 TCTCTCTCTGTGTCAGCATCCAAGGAGCT---TTCCCGTTGTCTGGTGAAAGGCAGCCTG 706
              |||||||| ||| |||||   ||  |||    |||   ||| ||| ||   |||||  ||
Db        466 TCTCTCTCCGTGGCAGCAGAAAATCAGCCATATTCTTATTGCCTGCTGTCTGGCAGTTTG 525

Qy        707 GGAATGAACATTGTTAGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGAT 766
              ||  ||||||| || ||| | ||||   |     ||||||| || ||  |  |    |||
Db        526 GGCTTGAACATCGTCAGTGCAATCTGCTCTGCAGTTGGAGTCATACTCTTCATCACAGAT 585

Qy        767 ATGTGCAT---------------CAATGGGGTAGCTGGCCAAGACTACTGGGCCGTGCTT 811
               |  | ||               | |    |    | . |   ||  |||||  |||
Db        586 CTAAGTATTCCCCACCCATATGCCTACCCCGACTATTATCCTTACGCCTGGGGTGTGAAC 645

Qy        812 TCTGGAAAAGGCATTTCAGCCACGCTGATGATCTTCTCCCTCTTGGAGTTCTTCGTAGCT 871
              ||||||  |  ||||| | |  |||| || |||||| |||| |||||||   | | ||
Db        646 CCTGGAATGGCGATTTCTGGCGTGCTGCTGGTCTTCTGCCTCCTGGAGTTTGGCATCGCA 705

Qy        872 TGTGCCACAGCCCATTTTGCCAACCA 897
              || ||  |  |||| |||| |  |||
Db        706 TGCGCATCTTCCCACTTTGGCTGCCA 731


Search completed: July 16, 2003, 13:27:52
Job time : 2399.68 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 16, 2003, 03:41:41 ; Search time 224.967 Seconds
(without alignments)
13183.634 Million cell updates/sec

Title: US-09-981-353-81
Perfect score: 1317
Sequence: 1 aaggaacaaagtaagtccat.....tttttccttaataaaatgtc 1317

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1317	100.0	1317	20	AAZ06368	Human receptor pro
	2	1108	84.1	1141	21	AAZ95381	Human colon specif
	3	1084.4	82.3	1291	21	AAA26287	Human secreted pro
	4	1079	81.9	1292	21	AAA26331	Human secreted pro
	5	712.8	54.1	854	23	AAS81191	DNA encoding novel
	6	389.6	29.6	441	21	AAC98717	Human colon cancer
	7	303.6	23.1	351	22	AAS37436	Novel human diagno
	8	292.6	22.2	342	22	AAS37580	Novel human diagno
	9	291.2	22.1	373	22	AAF66008	Novel human polynu
	10	288.6	21.9	383	22	AAS37423	Novel human diagno
	11	276	21.0	370	22	AAF65997	Novel human polynu
	12	81.2	6.2	1236	19	AAV48114	Nucleotide sequenc
	13	81.2	6.2	1336	21	AAZ51573	Human cell surface
	14	81.2	6.2	1336	24	ABK39631	cDNA encoding L985
	15	81.2	6.2	1374	21	AAC99077	Human pancreatic c
	16	81.2	6.2	1375	24	ABL90599	Human polynucleoti
	17	79.6	6.0	861	24	ABK39776	DNA encoding Ral2S
	18	79.6	6.0	1155	24	ABK39775	DNA encoding Ral2-
	19	79.6	6.0	1353	24	ABK39774	cDNA encoding lung
	20	74.8	5.7	932	22	AAD19723	Dendritic cell (DC
	21	74.8	5.7	945	19	AAV59668	Human secreted pro
	22	74.8	5.7	1001	21	AAZ52580	Human secreted pro
	23	74.8	5.7	1036	22	AAD19722	Dendritic cell (DC
	24	74.8	5.7	1106	22	ABA09233	Human secreted pro
	25	74.8	5.7	1106	22	AAH99750	Human protein enco
	26	74.8	5.7	1108	22	AAF63724	Human hHAIERbs-iso
	27	74.8	5.7	1289	22	AAH79291	Human cell surface
	28	74.8	5.7	1488	20	AAZ33586	Human breast tumou
	29	74.8	5.7	1522	20	AAZ06369	Human receptor pro
	30	74.8	5.7	1570	21	AAA26426	Human secreted pro
	31	74.8	5.7	1728	24	ABK34885	Human cDNA encodin
	32	74.8	5.7	1743	21	AAF18295	Lung cancer associ
	33	74.8	5.7	1871	22	AAH35024	Human colon cancer
	34	74.8	5.7	1871	24	ABL90750	Human polynucleoti
	35	74.8	5.7	2192	21	AAA27130	Human inflammation
	36	74.4	5.6	648	24	ABQ58726	Human colon cancer
	37	72	5.5	1060	20	AAX97730	Extended human sec
	38	71.2	5.4	2545	11	AAQ04645	Encodes beta subun
	39	71.2	5.4	2545	12	AAQ14734	Encodes beta subun
	40	71.2	5.4	2545	14	AAQ51021	Human FceRI beta g
	41	69.2	5.3	708	22	AAF77694	Murine wild-type F
	42	61.2	4.6	999	22	AAH99112	Human EST-derived
	43	61.2	4.6	1025	23	AAS76894	DNA encoding novel
	44	60	4.6	60	24	ABN38453	Human spliced tran
c	45	60	4.6	489	24	ABL38455	Human colon tumour

ALIGNMENTS

RESULT 1
AAZ06368
ID AAZ06368 standard; DNA; 1317 BP.
XX
AC AAZ06368;
XX
DT 26-OCT-1999 (first entry)
XX
DE Human receptor protein (HURP) 3 nucleotide sequence.
XX
KW receptor; cancer; autoimmune disorder; inflammation;
KW antagonist; cell surface protein; cell signalling;
KW antibody; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 221..1024
FT /*tag= a
FT /product= Human receptor protein 3
XX
PN WO9941375-A2.
XX
PD 19-AUG-1999.
XX
PF 05-FEB-1999; 99WO-US02572.
XX
PR 12-FEB-1998; 98US-0022939.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Au-Young J, Bandman O, Baughn M, Corley NC, Guegler KJ;

```
PI   Hillman JL,  Lal P,  Shah P,  Tang YT,  Yue H;
XX
DR   WPI; 1999-494536/41.
DR   P-PSDB; AAY15224.
XX
PT   New human receptor proteins, used e.g. to treat, prevent and
PT   diagnose gastrointestinal and developmental disorders - and related
PT   nucleic acids, vectors, transformed cells, antibodies, agonists and
PT   antagonists
XX
PS   Claim 11; Page 88; 94pp; English.
XX
CC   This is the consensus sequence of the Human receptor protein 3 (HURP-3)
CC   derived from the following overlapping and/or extended nucleic acid
CC   sequences; Incyte clones 611082 (COLNNOT01), 776770 (COLNNOT05), 904152
CC   (COLNNOT07), 908496 (COLNNOT09), and 1866665 (SKINBIT01).
CC   HURP-3 is expressed in cancerous, inflamed, fetal/proliferating, and
CC   reproductive tissue. HURP-3 therefore appears to have a role in some
CC   forms of cancer, autoimmune/inflammatory disorders, reproductive
CC   disorders, and developmental disorders.
CC   This gives rise to the possibility of using an antagonist or an antibody
CC   of HURP to treat or prevent cancer or autoimmune/inflammatory disorders.
XX
SQ   Sequence 1317 BP; 375 A; 280 C; 278 G; 384 T; 0 other;

  Query Match             100.0%;  Score 1317;  DB 20;  Length 1317;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1317;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGAACAAAGTAAGTCCATTGATACGTTCTTGCCTATCTCTCCTCCAAATCAATGGGCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGGAACAAAGTAAGTCCATTGATACGTTCTTGCCTATCTCTCCTCCAAATCAATGGGCA 60

Qy         61 CAAACTGTGGCTGGTCTACCTGTGTGGGTTCTGTTCTCTAGATTGGAGGGATGAAGACAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAACTGTGGCTGGTCTACCTGTGTGGGTTCTGTTCTCTAGATTGGAGGGATGAAGACAA 120

Qy        121 GTTCTTGACTCTATGTTGAGGCCAGTTGAAAAATGAGGGAGAATAAAACCATGAACGAAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTTCTTGACTCTATGTTGAGGCCAGTTGAAAAATGAGGGAGAATAAAACCATGAACGAAA 180

Qy        181 CAAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATCCAAGCCAAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATCCAAGCCAAC 240

Qy        241 AAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAGCTTTATGGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAGCTTTATGGC 300

Qy        301 TCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCAGGGTGCTCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCAGGGTGCTCA 360

Qy        361 GCGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCAACCGGGTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCAACCGGGTCA 420

Qy        421 AGGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTTTAAAGAAGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTTTAAAGAAGA 480

Qy        481 AGCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGGTTTTGGAAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AGCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGGTTTTGGAAT 540

Qy        541 TGTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCTACTGCTGT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TGTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCTACTGCTGT 600

Qy        601 TATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCTCTCTCTGT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCTCTCTCTGT 660

Qy        661 GTCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATGAACATTGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GTCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATGAACATTGT 720

Qy        721 TAGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGCATCAATGG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TAGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGCATCAATGG 780

Qy        781 GGTAGCTGGCCAAGACTACTGGGCCGTGCTTTCTGGAAAAGGCATTTCAGCCACGCTGAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GGTAGCTGGCCAAGACTACTGGGCCGTGCTTTCTGGAAAAGGCATTTCAGCCACGCTGAT 840

Qy        841 GATCTTCTCCCTCTTGGAGTTCTTCGTAGCTTGTGCCACAGCCCATTTTGCCAACCAAGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GATCTTCTCCCTCTTGGAGTTCTTCGTAGCTTGTGCCACAGCCCATTTTGCCAACCAAGC 900

Qy        901 AAACACCACAACCAATATGTCTGTCCTGGTTATTCCAAATATGTATGAAAGCAACCCTGT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AAACACCACAACCAATATGTCTGTCCTGGTTATTCCAAATATGTATGAAAGCAACCCTGT 960

Qy        961 GACACCAGCGTCTTCTTCAGCTCCTCCCAGATGCAACAACTACTCAGCTAATGCCCCTAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GACACCAGCGTCTTCTTCAGCTCCTCCCAGATGCAACAACTACTCAGCTAATGCCCCTAA 1020

Qy       1021 ATAGTAAAAGAAAAAGGGGTATCAGTCTAATCTCATGGAGAAAAACTACTTGCAAAAACT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATAGTAAAAGAAAAAGGGGTATCAGTCTAATCTCATGGAGAAAAACTACTTGCAAAAACT 1080

Qy       1081 TCTTAAGAAGATGTCTTTTATTGTCTACAATGATTTCTAGTCTTTAAAAACTGTGTTTGA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TCTTAAGAAGATGTCTTTTATTGTCTACAATGATTTCTAGTCTTTAAAAACTGTGTTTGA 1140

Qy       1141 GATTTGTTTTTAGGTTGGTCGCTAATGATGGCTGTATCTCCCTTCACTGTCTCTTCCTAC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GATTTGTTTTTAGGTTGGTCGCTAATGATGGCTGTATCTCCCTTCACTGTCTCTTCCTAC 1200

Qy       1201 ATTACCACTACTACATGCTGGCAAAGGTGAAGGATCAGAGGACTGAAAAATGATTCTGCA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 ATTACCACTACTACATGCTGGCAAAGGTGAAGGATCAGAGGACTGAAAAATGATTCTGCA 1260

Qy       1261 ACTCTCTTAAAGTTAGAAATGTTTCTGTTCATATTACTTTTTCCTTAATAAAATGTC 1317
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 ACTCTCTTAAAGTTAGAAATGTTTCTGTTCATATTACTTTTTCCTTAATAAAATGTC 1317


RESULT 2
AAZ95381
ID   AAZ95381 standard; DNA; 1141 BP.
XX
AC   AAZ95381;
XX
DT   01-JUN-2000  (first entry)
XX
DE   Human colon specific gene (CSG) SEQ ID NO:3.
XX
KW   Human; colon specific gene; CSG; diagnosis; colon cancer; detection;
KW   cytostatic; ds.
XX
OS   Homo sapiens.
XX
PN   WO200007632-A1.
XX
PD   17-FEB-2000.
XX
PF   20-JUL-1999;   99WO-US16357.
XX
PR   04-AUG-1998;   98US-0095231.
XX
PA   (DIAD-) DIADEXUS LLC.
XX
PI   Sun Y,  Recipon H,  Macina RA;
XX
DR   WPI; 2000-205579/18.
```

```
XX
PT   Novel methods for diagnosing, monitoring, staging, imaging and treating
PT   colon cancer by measuring the level of colon specific gene markers -
XX
PS   Claim 6; Page 38-39; 42pp; English.
XX
CC   The present invention describes a method for diagnosing the presence of
CC   colon cancer in a patient. The method comprises measuring levels of
CC   colon specific gene markers (CSG) in cells, tissues or bodily fluids,
CC   and comparing the measured levels of CSG with levels of CSG from a
CC   normal human control, where an increase in measured CSG levels in the
CC   patient versus control is associated with the presence of colon cancer.
CC   AAZ95379 to AAZ95381 represent human CSG sequences. The method is used
CC   to detect, monitor, stage or give a prognosis for colon cancer.
CC   Antibodies against CSGs are used for detection or image localisation of
CC   the CSGs. The antibodies can be conjugated to cytotoxic agent or drug and
CC   used to treat colon cancer. The method of the invention is more accurate
CC   than prior art clinical methods for staging colon cancer, because it
CC   measures colon specific markers, and, unlike pathological staging
CC   methods, do not depend on an invasive procedure.
XX
SQ   Sequence 1141 BP; 317 A; 248 C; 238 G; 338 T; 0 other;

  Query Match             84.1%;  Score 1108;  DB 21;  Length 1141;
  Best Local Similarity   98.2%;  Pred. No. 2.4e-285;
  Matches 1120;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy        178 AAACAAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATCCAAGCC 237
              | | ||  | ||||| |   |      |   | ||||||||||||||||||||||||||
Db          2 AGAGAAAGAGGAAACATAGAGGTGCCAAAGGAACAAAGACATAATGATGTCATCCAAGCC 61

Qy        238 AACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAGCTTTAT 297
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 AACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAGCTTTAT 121

Qy        298 GGCTCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCAGGGTGC 357
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 GGCTCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCAGGGTGC 181

Qy        358 TCAGCGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCAACCGGG 417
              ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        182 TCAGCGTGCTCAGCCCTATGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCAACCGGG 241

Qy        418 TCAAGGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTTTAAAGA 477
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 TCAAGGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTTTAAAGA 301

Qy        478 AGAAGCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGGTTTTGG 537
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 AGAAGCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGGTTTTGG 361

Qy        538 AATTGTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCTACTGC 597
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 AATTGTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCTACTGC 421

Qy        598 TGTTATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCTCTCTC 657
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        422 TGTTATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCTCTCTC 481

Qy        658 TGTGTCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATGAACAT 717
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        482 TGTGTCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATGAACAT 541

Qy        718 TGTTAGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGCATCAA 777
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        542 TGTTAGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGCATCAA 601

Qy        778 TGGGGTAGCTGGCCAAGACTACTGGGCCGTGCTTTCTGGAAAAGGCATTTCAGCCACGCT 837
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        602 TGGGGTAGCTGGCCAAGACTACTGGGCCGTGCTTTCTGGAAAAGGCATTTCAGCCACGCT 661

Qy        838 GATGATCTTCTCCCTCTTGGAGTTCTTCGTAGCTTGTGCCACAGCCCATTTTGCCAACCA 897
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        662 GATGATCTTCTCCCTCTTGGAGTTCTTCGTAGCTTGTGCCACAGCCCATTTTGCCAACCA 721

Qy        898 AGCAAACACCACAACCAATATGTCTGTCCTGGTTATTCCAAATATGTATGAAAGCAACCC 957
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        722 AGCAAACACCACAACCAATATGTCTGTCCTGGTTATTCCAAATATGTATGAAAGCAACCC 781

Qy        958 TGTGACACCAGCGTCTTCTTCAGCTCCTCCCAGATGCAACAACTACTCAGCTAATGCCCC 1017
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        782 TGTGACACCAGCGTCTTCTTCAGCTCCTCCCAGATGCAACAACTACTCAGCTAATGCCCC 841

Qy       1018 TAAATAGTAAAAGAAAAAGGGGTATCAGTCTAATCTCATGGAGAAAAACTACTTGCAAAA 1077
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        842 TAAATAGTAAAAGAAAAAGGGGTATCAGTCTAATCTCATGGAGAAAAACTACTTGCAAAA 901

Qy       1078 ACTTCTTAAGAAGATGTCTTTTATTGTCTACAATGATTTCTAGTCTTTAAAAACTGTGTT 1137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        902 ACTTCTTAAGAAGATGTCTTTTATTGTCTACAATGATTTCTAGTCTTTAAAAACTGTGTT 961

Qy       1138 TGAGATTTGTTTTTAGGTTGGTCGCTAATGATGGCTGTATCTCCCTTCACTGTCTCTTCC 1197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        962 TGAGATTTGTTTTTAGGTTGGTCGCTAATGATGGCTGTATCTCCCTTCACTGTCTCTTCC 1021

Qy       1198 TACATTACCACTACTACATGCTGGCAAAGGTGAAGGATCAGAGGACTGAAAAATGATTCT 1257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1022 TACATTACCACTACTACATGCTGGCAAAGGTGAAGGATCAGAGGACTGAAAAATGATTCT 1081

Qy       1258 GCAACTCTCTTAAAGTTAGAAATGTTTCTGTTCATATTACTTTTTCCTTAATAAAATGTC 1317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1082 GCAACTCTCTTAAAGTTAGAAATGTTTCTGTTCATATTACTTTTTCCTTAATAAAATGTC 1141


RESULT 3
AAA26287
ID   AAA26287 standard; cDNA; 1291 BP.
XX
AC   AAA26287;
XX
DT   29-JUN-2000  (first entry)
XX
DE   Human secreted protein gene 7 SEQ ID NO:17.
XX
KW   Human; secreted protein; diagnosis; neuroprotective; nootropic;
KW   neuroleptic; antimanic; cerebroprotective; immunomodulatory;
KW   anti-microbial; cardiant; cytostatic; antiinflammatory; haemostatic;
KW   anticonvulsant; vasotropic; vaccine; gene therapy; anti-sense therapy;
KW   neural; reproductive; immune disorder; immunodeficiency; infection;
KW   lymphoma; demyelinating disease; autoimmunity; cancer; inflammation;
KW   aneurysm; haemorrhage; Alzheimer's disease; Parkinson's disease;
KW   Huntington's disease; Tourette syndrome; multiple sclerosis; meningitis;
KW   ischaemia; mania; dementia; obsessive compulsive disorder;
KW   viral prophylaxis; developmental disorder; sexually-linked disorder;
KW   cardiovascular disorder; food additive; preservative; ss.
XX
OS   Homo sapiens.
XX
PN   WO200011014-A1.
XX
PD   02-MAR-2000.
XX
PF   24-AUG-1999;   99WO-US19330.
XX
PR   25-AUG-1998;   98US-0097917.
PR   31-AUG-1998;   98US-0098634.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Moore PA,  Ruben SM,  Olsen HS,  Shi Y,  Rosen CA,  Florence KA;
PI   Soppet DR,  Lafleur DW,  Endress GA,  Ebner R,  Komatsoulis G;
PI   Duan RD;
XX
DR   WPI; 2000-224656/19.
DR   P-PSDB; AAY91352.
XX
```

```
PT   Novel secreted proteins and corresponding DNA molecules that can be
PT   used to prevent, treat and diagnose disease in humans, for example,
PT   Alzheimer's, cancer, and immune disorders -
XX
PS   Claim 1; Page 315-316; 416pp; English.
XX
CC   The polynucleotide sequences given in AAA26281 to AAA26336 encode the
CC   human secreted proteins given in AAY91346 to AAY91449. The human secreted
CC   proteins can have activities based on the tissues and cells they are
CC   expressed in. Examples of the activities are: neuroprotective; nootropic;
CC   neuroleptic; antimanic; cerebroprotective; immunomodulatory;
CC   anti-microbial; cardiant; cytostatic; antiinflammatory; haemostatic;
CC   anticonvulsant; and vasotropic. The polynucleotides and proteins may be
CC   used to prevent, treat or ameliorate a medical condition, e.g. by protein
CC   or gene therapy. Conditions treatable by the proteins of the invention
CC   include neural, reproductive, or immune disorders, especially
CC   immunodeficiency, infection, lymphomas, demyelinating diseases,
CC   auto-immunities, cancer, general microbial infection, inflammation,
CC   aneurysms and haemorrhages. Specific examples include: Alzheimer's
CC   disease; Parkinson's; Huntington's; Tourette syndrome; multiple
CC   sclerosis; meningitis; ischaemia; prostate cancer; mania; dementia;
CC   obsessive compulsive disorder and viral prophylaxis. The polynucleotides
CC   and proteins can also be used in the detection of disorders associated
CC   with the function of the protein, for example, the detection of
CC   developmental disorders, sexually-linked disorders, or disorders of the
CC   cardiovascular system. They may also be used as food additives or
CC   preservatives. AAA26272 to AAA26280 and AAY91345 are sequences used in
CC   the exemplification of the present invention.
XX
SQ   Sequence 1291 BP; 371 A; 285 C; 272 G; 358 T; 5 other;

  Query Match             82.3%;  Score 1084.4;  DB 21;  Length 1291;
  Best Local Similarity   97.8%;  Pred. No. 5.1e-279;
  Matches 1111;  Conservative    1;  Mismatches   17;  Indels    7;  Gaps    1;

Qy        182 AAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATCCAAGCCAACA 241
              ||  | ||||||  |   |      |   | ||||||||||||||||||||||||||||||
Db         75 AAAGAGGAAACATAGAGGTGCCAAAGGAACAAAGACATAATGATGTCATCCAAGCCAACA 134

Qy        242 AGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAGCTTTATGGCT 301
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        135 AGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAGCTTTATGGCT 194

Qy        302 CCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCAGGGTGCTCAG 361
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        195 CCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCAGGGTGCTCAG 254

Qy        362 CGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCAACCGGGTCAA 421
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        255 CGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCAACCGGGTCAA 314

Qy        422 GGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTTTAAAGAAGAA 481
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        315 GGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTTTAAAGAAGAA 374

Qy        482 GCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGGTTTTGGAATT 541
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        375 GCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGGTTTTGGAATT 434

Qy        542 GTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCTACTGCTGTT 601
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        435 GTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCTACTGCTGTT 494

Qy        602 ATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCTCTCTCTGTG 661
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        495 ATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCTCTCTCTGTG 554

Qy        662 TCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATGAACATTGTT 721
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db        555 TCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATGAACATTGKT 614

Qy        722 AGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGCATCAATGGG 781
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        615 AGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGCATCAATGGG 674

Qy        782 GTAGCTGGCCAAGACTACTGGGCCGTGCTTTCTGGAAAAGGCATTTCAGCCACGCTGATG 841
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        675 GTAGCTGGCCAAGACTACTGGGCCGTGCTTTCTGGAAAAGGCATTTCAGCCACGCTGATG 734

Qy        842 ATCTTCTCCCTCTTGGAGTTCTTCGTAGCTTGTGCCACAGCCCATTTTGCCAACCAAGCA 901
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        735 ATCTTCTCCCTCTTGGAGTTCTTCGTAGCTTGTGCCACAGCCCATTTTGCCAACCAAGCA 794

Qy        902 AACACCACAACCAATATGTCTGTCCTGGTTATTCCAAATATGTATGAAAGCAACCCTGTG 961
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        795 AACACCACAACCAATATGTCTGTCCTGGTTATTCCAAATATGTATGAAAGCAACCCTGTG 854

Qy        962 ACACCAGCGTCTTCTTCAGCTCCTCCCAGATGCAACAACTACTCAGCTAATGCCCCTAAA 1021
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        855 ACACCAGCGTCTTCTTCAGCTCCTCCCAGATGCAACAACTACTCAGCTAATGCCCC---- 910

Qy       1022 TAGTAAAAGAAAAAGGGGTATCAGTCTAATCTCATGGAGAAAAACTACTTGCAAAAACTT 1081
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        911 ---TAAAAGAAAAAGGGGTATCAGTCTAATCTCATGGAGAAAAACTACTTGCAAAAACTT 967

Qy       1082 CTTAAGAAGATGTCTTTTATTGTCTACAATGATTTCTAGTCTTTAAAAACTGTGTTTGAG 1141
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        968 CTTAAGAAGATGTCTTTTATTGTCTACAATGATTTCTAGTCTTTAAAAACTGTGTTTGAG 1027

Qy       1142 ATTTGTTTTTAGGTTGGTCGCTAATGATGGCTGTATCTCCCTTCACTGTCTCTTCCTACA 1201
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1028 ATTTGTTTTTAGGTTGGTCGCTAATGATGGCTGTATCTCCCTTCACTGTCTCTTCCTACA 1087

Qy       1202 TTACCACTACTACATGCTGGCAAAGGTGAAGGATCAGAGGACTGAAAAATGATTCTGCAA 1261
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1088 TTACCACTACTACATGCTGGCAAAGGTGAAGGATCAGAGGACTGAAAAATGATTCTGCAA 1147

Qy       1262 CTCTCTTAAAGTTAGAAATGTTTCTGTTCATATTACTTTTTCCTTAATAAAATGTC 1317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1148 CTCTCTTAAAGTTAGAAATGTTTCTGTTCATATTACTTTTTCCTTAATAAAATGTC 1203


RESULT 4
AAA26331
ID   AAA26331 standard; cDNA; 1292 BP.
XX
AC   AAA26331;
XX
DT   29-JUN-2000  (first entry)
XX
DE   Human secreted protein gene 7 SEQ ID NO:61.
XX
KW   Human; secreted protein; diagnosis; neuroprotective; nootropic;
KW   neuroleptic; antimanic; cerebroprotective; immunomodulatory;
KW   anti-microbial; cardiant; cytostatic; antiinflammatory; haemostatic;
KW   anticonvulsant; vasotropic; vaccine; gene therapy; anti-sense therapy;
KW   neural; reproductive; immune disorder; immunodeficiency; infection;
KW   lymphoma; demyelinating disease; autoimmunity; cancer; inflammation;
KW   aneurysm; haemorrhage; Alzheimer's disease; Parkinson's disease;
KW   Huntington's disease; Tourette syndrome; multiple sclerosis; meningitis;
KW   ischaemia; mania; dementia; obsessive compulsive disorder;
KW   viral prophylaxis; developmental disorder; sexually-linked disorder;
KW   cardiovascular disorder; food additive; preservative; ss.
XX
OS   Homo sapiens.
XX
PN   WO200011014-A1.
XX
PD   02-MAR-2000.
XX
PF   24-AUG-1999;   99WO-US19330.
XX
PR   25-AUG-1998;   98US-0097917.
PR   31-AUG-1998;   98US-0098634.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
```

```
XX
PI   Moore PA,  Ruben SM,  Olsen HS,  Shi Y,  Rosen CA,  Florence KA;
PI   Soppet DR,  Lafleur DW,  Endress GA,  Ebner R,  Komatsoulis G;
PI   Duan RD;
XX
DR   WPI; 2000-224656/19.
DR   P-PSDB; AAY91396.
XX
PT   Novel secreted proteins and corresponding DNA molecules that can be
PT   used to prevent, treat and diagnose disease in humans, for example,
PT   Alzheimer's, cancer, and immune disorders -
XX
PS   Claim 1; Page 350-351; 416pp; English.
XX
CC   The polynucleotide sequences given in AAA26281 to AAA26336 encode the
CC   human secreted proteins given in AAY91346 to AAY91449. The human secreted
CC   proteins can have activities based on the tissues and cells they are
CC   expressed in. Examples of the activities are: neuroprotective; nootropic;
CC   neuroleptic; antimanic; cerebroprotective; immunomodulatory;
CC   anti-microbial; cardiant; cytostatic; antiinflammatory; haemostatic;
CC   anticonvulsant; and vasotropic. The polynucleotides and protein's may be
CC   used to prevent, treat or ameliorate a medical condition, e.g. by protein
CC   or gene therapy. Conditions treatable by the proteins of the invention
CC   include neural, reproductive, or immune disorders, especially
CC   immunodeficiency, infection, lymphomas, demyelinating diseases,
CC   auto-immunities, cancer, general microbial infection, inflammation,
CC   aneurysms and haemorrhages. Specific examples include: Alzheimer's
CC   disease; Parkinson's; Huntington's; Tourette syndrome; multiple
CC   sclerosis; meningitis; ischaemia; prostate cancer; mania; dementia;
CC   obsessive compulsive disorder and viral prophylaxis. The polynucleotides
CC   and proteins can also be used in the detection of disorders associated
CC   with the function of the protein, for example, the detection of
CC   developmental disorders, sexually-linked disorders, or disorders of the
CC   cardiovascular system. They may also be used as food additives or
CC   preservatives. AAA26272 to AAA26280 and AAY91345 are sequences used in
CC   the exemplification of the present invention.
XX
SQ   Sequence 1292 BP; 372 A; 284 C; 272 G; 355 T; 9 other;

  Query Match            81.9%;  Score 1079;  DB 21;  Length 1292;
  Best Local Similarity  97.3%;  Pred. No. 1.4e-277;
  Matches 1105;  Conservative    4;  Mismatches   20;  Indels    7;  Gaps    1;

Qy        182 AAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATCCAAGCCAACA 241
              ||  | ||||||  |   |       |   | ||||||||||||||||||||||||||||||
Db         76 AAAGAGGAAACATAGAGGTGCCAAAGGAACAAAGACATAATGATGTCATCCAAGCCAACA 135

Qy        242 AGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAGCTTTATGGCT 301
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 AGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAGCTTTATGGCT 195

Qy        302 CCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCAGGGTGCTCAG 361
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 CCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCAGGGTGCTCAG 255

Qy        362 CGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCAACCGGGTCAA 421
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 CGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCAACCGGGTCAA 315

Qy        422 GGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTTTAAAGAAGAA 481
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 GGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTTTAAAGAAGAA 375

Qy        482 GCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGGTTTTGGAATT 541
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 GCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGGTTTTGGAATT 435

Qy        542 GTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCTACTGCTGTT 601
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db        436 GTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCTACTGCTGKT 495

Qy        602 ATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCTCTCTCTGTG 661
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 ATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCTCTCTCTGTG 555

Qy        662 TCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATGAACATTGTT 721
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        556 TCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATGAACATTGGT 615

Qy        722 AGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGCATCAATGGG 781
              || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 AGGTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGCATCAATGGG 675

Qy        782 GTAGCTGGCCAAGACTACTGGGCCGTGCTTTCTGGAAAAGGCATTTCAGCCACGCTGATG 841
              |||:|||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        676 GTARCTGGCCAAGACTACTGGNCCGTGCTTTCTGGAAAAGGCATTTCAGCCACGCTGATG 735

Qy        842 ATCTTCTCCCTCTTGGAGTTCTTCGTAGCTTGTGCCACAGCCCATTTTGCCAACCAAGCA 901
              |||||||||:||||||||||||||||||||||||||||||||||||||||||||||||||
Db        736 ATCTTCTCCYTCTTGGAGTTCTTCGTAGCTTGTGCCACAGCCCATTTTGCCAACCAAGCA 795

Qy        902 AACACCACAACCAATATGTCTGTCCTGGTTATTCCAAATATGTATGAAAGCAACCCTGTG 961
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db        796 AACACCACAACCAATATGTCTGTCCTGGTTATTCCAAATATGTATGAAAGCAACCCTGKG 855

Qy        962 ACACCAGCGTCTTCTTCAGCTCCTCCCAGATGCAACAACTACTCAGCTAATGCCCCTAAA 1021
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        856 ACACCAGCGTCTTCTTCAGCTCCTCCCAGATGCAACAACTACTCAGCTAATGCCCC---- 911

Qy       1022 TAGTAAAAGAAAAAGGGGTATCAGTCTAATCTCATGGAGAAAAACTACTTGCAAAAACTT 1081
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        912 ---TAAAAGAAAAAGGGGTATCAGTCTAATCTCATGGAGAAAAACTACTTGCAAAAACTT 968

Qy       1082 CTTAAGAAGATGTCTTTTATTGTCTACAATGATTTCTAGTCTTTAAAAACTGTGTTTGAG 1141
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        969 CTTAAGAAGATGTCTTTTATTGTCTACAATGATTTCTAGTCTTTAAAAACTGTGTTTGAG 1028

Qy       1142 ATTTGTTTTTAGGTTGGTCGCTAATGATGGCTGTATCTCCCTTCACTGTCTCTTCCTACA 1201
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1029 ATTTGTTTTTAGGTTGGTCGCTAATGATGGCTGTATCTCCCTTCACTGTCTCTTCCTACA 1088

Qy       1202 TTACCACTACTACATGCTGGCAAAGGTGAAGGATCAGAGGACTGAAAAATGATTCTGCAA 1261
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1089 TTACCACTACTACATGCTGGCAAAGGTGAAGGATCAGAGGACTGAAAAATGATTCTGCAA 1148

Qy       1262 CTCTCTTAAAGTTAGAAATGTTTCTGTTCATATTACTTTTTCCTTAATAAAATGTC 1317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1149 CTCTCTTAAAGTTAGAAATGTTTCTGTTCATATTACTTTTTCCTTAATAAAATGTC 1204


RESULT 5
AAS81191
ID   AAS81191 standard; cDNA; 854 BP.
XX
AC   AAS81191;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #16995.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
```

```
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG17004.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 16995; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 854 BP; 234 A; 192 C; 197 G; 231 T; 0 other;

  Query Match            54.1%;  Score 712.8;  DB 23;  Length 854;
  Best Local Similarity.  97.1%;  Pred. No. 5.9e-180;
  Matches  726;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy        172 TGAACGAAACAAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATC 231
              || |  | | ||  | ||||||  |   |      |   | ||||||||||||||||||
Db          8 TGGAGCAGAGAAAGAGGAAACATAGAGGTGCCAAAGGAACAAAGACATAATGATGTCATC 67

Qy        232 CAAGCCAACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAG 291
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         68 CAAGCCAACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAG 127

Qy        292 CTTTATGGCTCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCA 351
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        128 CTTTATGGCTCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCA 187

Qy        352 GGGTGCTCAGCGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCA 411
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        188 GGGTGCTCAGCGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCA 247

Qy        412 ACCGGGTCAAGGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTT 471
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        248 ACCGGGTCAAGGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTT 307

Qy        472 TAAAGAAGAAGCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGG 531
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        308 TAAAGAAGAAGCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGG 367

Qy        532 TTTTGGAATTGTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTC 591
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        368 TTTTGGAATTGTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTC 427

Qy        592 TACTGCTGTTATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTC 651
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        428 TACTGCTGTTATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTC 487

Qy        652 TCTCTCTGTGTCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAAT 711
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        488 TCTCTCTGTGTCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAAT 547

Qy        712 GAACATTGTTAGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTG 771
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        548 GAACATTGTTAGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTG 607

Qy        772 CATCAATGGGGTAGCTGGCCAAGACTACTGGGCCGTGCTTTCTGGAAAAGGCATTTCAGC 831
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        608 CATCAATGGGGTAGCTGGCCAAGACTACTGGGCCGTGCTTTCTGGAAAAGGCATTTCAGC 667

Qy        832 CACGCTGATGATCTTCTCCCTCTTGGAGTTCTTCGTAGCTTGTGCCACAGCCCATTTTGC 891
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        668 CACGCTGATGATCTTCTCCCTCTTGGAGTTCTTCGTAGCTTGTGCCACAGCCCATTTTGC 727

Qy        892 CAACCAAGCAAACACCACAACCAATATG 919
              |||||||||||||||||||||||||||||
Db        728 CAACCAAGCAAACACCACAACCAATATG 755


RESULT 6
AAC98717
ID   AAC98717 standard; cDNA; 441 BP.
XX
AC   AAC98717;
XX
DT   09-MAR-2001  (first entry)
XX
DE   Human colon cancer antigen nucleotide sequence SEQ ID NO:727.
XX
KW   Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW   identification; cytostatic; cardioactive; neuroprotective; vulnerary;
KW   immunomodulatory; muscular; gynaecological; gastrointestinal;
KW   nephrotropic; antiinfective; antibacterial; gene therapy; wound;
KW   neural disorder; immune system disorder; muscular disorder;
KW   reproductive disorder; gastrointestinal disorder; renal disorder;
KW   infectious disease; cardiovascular disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055351-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05883.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-587534/55.
DR   P-PSDB; AAB53960.
XX
PT   Colon cancer associated gene sequences, referred to as colon cancer
PT   antigens, useful for the treatment, prevention, and diagnosis of colon
PT   disorders such as colon cancer -
XX
PS   Claim 1; Page 1284; 2104pp; English.
XX
CC   AAC97991 to AAC98763 encode the human colon cancer associated proteins,
CC   called human colon cancer antigens, given in AAB53234 to AAB54006. The
CC   human colon cancer antigens can have cytostatic, cardioactive, muscular;
CC   neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
CC   vulnerary, nephrotropic, antiinfective and antibacterial activities, and
CC   can be used in gene therapy. The colon cancer antigen polynucleotides,
CC   proteins and antibodies to the proteins are useful for the prevention,
CC   treatment and diagnosis of colon disorders, such as colon cancer. The
CC   polynucleotides may be used in diagnostics and research, such as for
CC   chromosome identification, and as hybridisation probes. The proteins
CC   may also be used to prevent diseases such as neural disorders, immune
CC   system disorders, muscular disorders, reproductive disorders,
```

```
CC   gastrointestinal disorders, wounds, renal disorders, infectious
CC   diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
CC   AAB54007 represent sequences used in the exemplification of the present
CC   invention.
XX
SQ   Sequence 441 BP; 156 A; 87 C; 100 G; 92 T; 6 other;

  Query Match             29.6%;  Score 389.6;  DB 21;  Length 441;
  Best Local Similarity   96.6%;  Pred. No. 7.5e-94;
  Matches  427;  Conservative    0;  Mismatches   10;  Indels    5;  Gaps    3;

Qy        113 GAAGACAAGTTCTTGACTCTATGTTGAGGCCAGTTGAAAAATGAGGGAGAATAAAACCAT 172
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 GAAGACAAGTTCTTGACTCTATGTTGAGGCCAGTTGAAAAATGAGGGAGAATAAAACCAT 61

Qy        173 GAACGAAACAAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATCC 232
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 GAACGAAACAAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATCC 121

Qy        233 AAGCCAACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAGC 292
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 AAGCCAACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAGC 181

Qy        293 TTTATGGCTCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCAG 352
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 TTTATGGCTCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCAG 241

Qy        353 GGTGCTCAGCGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCAA 412
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 GGTGCTCAGCGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCAA 301

Qy        413 CCGGGTCAAGGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTTT 472
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        302 CCGGGTCAAGGAAATATACNAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTTT 361

Qy        473 AAA--GAAGAAGCAAAGGCACTA-GGGGTGATCCAGATCATGGTTGGATTGATGCACATT 529
              |||   |||||||||||||||| |||||||| ||||||||| |||||||||  |   |
Db        362 AAAAGAAAGAAGCAAAGGCACTAGGGGGTGATNCAGATCATGGNTGGATTGATG--CCAT 419

Qy        530 GGTTTTGGAATTGTTTTGTGTT 551
                | |||||||||| |||| |||
Db        420 TGNTTTGGAATTGNTTTGNGTT 441


RESULT 7
AAS37436
ID   AAS37436 standard; cDNA; 351 BP.
XX
AC   AAS37436;
XX
DT   17-DEC-2001  (first entry)
XX
DE   Novel human diagnostic and therapeutic gene #494.
XX
KW   Human; cancer; breast; lung; colon; prostate; cytostatic; diagnostic; ss.
XX
OS   Homo sapiens.
XX
PN   WO200166753-A2.
XX
PD   13-SEP-2001.
XX
PF   09-MAR-2001; 2001WO-US07787.
XX
PR   09-MAR-2000; 2000US-0188609.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Williams LT,  Escobedo J,  Innis MA,  Garcia PD,  Sudduth-Klinger J;
PI   Reinhard C,  Randazzo F,  Kennedy GC,  Pot D,  Kassam A,  Lamson G;
PI   Drmanac R,  Crkvenjakov R,  Dickson M,  Drmanac S,  Labat I;
PI   Leshkowitz D,  Kita D,  Garcia V,  Jones WL,  Stache-Crain B;
XX
DR   WPI; 2001-530177/58.
XX
PT   New polynucleotides and polypeptides, useful for diagnosis and
PT   treatment of breast, lung and colon cancer -
XX
PS   Claim 1; Page 720; 1193pp; English.
XX
CC   The invention relates to new polynucleotides and polypeptides, useful for
CC   diagnosis and treatment of breast, lung and colon cancer. The sequences
CC   can be used in detecting differentially expressed genes correlated with a
CC   cancerous state of a mammalian cell, comprising detecting at least one
CC   differentially expressed gene product in a test sample derived from a
CC   cell suspected of being cancerous. They can also be used to inhibit
CC   tumour growth by modulating expression of a gene product. AAS36943-
CC   AAS39338 represent novel human diagnostic and therapeutic coding
CC   sequences of the invention.
XX
SQ   Sequence 351 BP; 123 A; 79 C; 81 G; 68 T; 0 other;

  Query Match             23.1%;  Score 303.6;  DB 22;  Length 351;
  Best Local Similarity   93.0%;  Pred. No. 6e-71;
  Matches  318;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy        172 TGAACGAAACAAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATC 231
              || |  | | ||  | |||||| |   |      |   | ||||||||||||||||||||
Db         10 TGGAGCAGAGAAAGAGGAAACATAGAGGTGCCAAAGGAACAAAGACATAATGATGTCATC 69

Qy        232 CAAGCCAACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAG 291
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         70 CAAGCCAACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAG 129

Qy        292 CTTTATGGCTCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCA 351
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        130 CTTTATGGCTCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCA 189

Qy        352 GGGTGCTCAGCGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCA 411
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        190 GGGTGCTCAGCGTGCTCAGCCCTATGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCA 249

Qy        412 ACCGGGTCAAGGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTT 471
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        250 ACCGGGTCAAGGAAATATATAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTT 309

Qy        472 TAAAGAAGAAGCAAAGGCACTAGGGGTGATCCAGATCATGGT 513
              |||||||||||||||||||||||||||||||||||||||||||
Db        310 TAAAGAAGAAGCAAAGGCACTAGGGGTGATCCAGATCATGGT 351


RESULT 8
AAS37580
ID   AAS37580 standard; cDNA; 342 BP.
XX
AC   AAS37580;
XX
DT   17-DEC-2001  (first entry)
XX
DE   Novel human diagnostic and therapeutic gene #638.
XX
KW   Human; cancer; breast; lung; colon; prostate; cytostatic; diagnostic; ss.
XX
OS   Homo sapiens.
XX
PN   WO200166753-A2.
XX
PD   13-SEP-2001.
XX
PF   09-MAR-2001; 2001WO-US07787.
XX
PR   09-MAR-2000; 2000US-0188609.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
```

```
XX
PI   Williams LT,  Escobedo J,  Innis MA,  Garcia PD,  Sudduth-Klinger J;
PI   Reinhard C,  Randazzo F,  Kennedy GC,  Pot D,  Kassam A,  Lamson G;
PI   Drmanac R,  Crkvenjakov R,  Dickson M,  Drmanac S,  Labat I;
PI   Leshkowitz D,  Kita D,  Garcia V,  Jones WL,  Stache-Crain B;
XX
DR   WPI; 2001-530177/58.
XX
PT   New polynucleotides and polypeptides, useful for diagnosis and
PT   treatment of breast, lung and colon cancer -
XX
PS   Claim 1; Page 756; 1193pp; English.
XX
CC   The invention relates to new polynucleotides and polypeptides, useful for
CC   diagnosis and treatment of breast, lung and colon cancer. The sequences
CC   can be used in detecting differentially expressed genes correlated with a
CC   cancerous state of a mammalian cell, comprising detecting at least one
CC   differentially expressed gene product in a test sample derived from a
CC   cell suspected of being cancerous. They can also be used to inhibit
CC   tumour growth by modulating expression of a gene product. AAS36943-
CC   AAS39338 represent novel human diagnostic and therapeutic coding
CC   sequences of the invention.
XX
SQ   Sequence 342 BP; 119 A; 78 C; 78 G; 65 T; 2 other;

  Query Match             22.2%;  Score 292.6;  DB 22;  Length 342;
  Best Local Similarity   92.5%;  Pred. No. 5.1e-68;
  Matches  307;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy        172 TGAACGAAACAAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATC 231
              || |  | | ||  | ||||||  |   |       |   | ||||||||||||||||||
Db         10 TGGAGCAGAGAAAGAGGAAACATAGAGGTGCCAAAGGAACAAAGACATAATGATGTCATC 69

Qy        232 CAAGCCAACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAG 291
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         70 CAAGCCAACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAG 129

Qy        292 CTTTATGGCTCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCA 351
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        130 CTTTATGGCTCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCA 189

Qy        352 GGGTGCTCAGCGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCA 411
              ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        190 GGGTGCTCAGCGTGCTCAGCCCTATGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCA 249

Qy        412 ACCGGGTCAAGGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTT 471
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        250 ACCGGGTCAAGGAAATATATAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTT 309

Qy        472 TAAAGAAGAAGCAAAGGCACTAGGGGTGATCC 503
              |||||||||||| ||||||||||||||||||
Db        310 TAAAGAAGAAGCANAGGCACTAGGGGTGATCC 341


RESULT 9
AAF66008
ID   AAF66008 standard; cDNA; 373 BP.
XX
AC   AAF66008;
XX
DT   09-APR-2001  (first entry)
XX
DE   Novel human polynucleotide, SEQ ID NO: 1764.
XX
KW   Human; cytostatic; gene therapy; colon cancer; prostate cancer;
KW   breast cancer; lung cancer; cancer detection; ss.
XX
OS   Homo sapiens.
XX
PN   WO200102568-A2.
XX
PD   11-JAN-2001.
XX
PF   30-JUN-2000; 2000WO-US18374.
XX
PR   02-JUL-1999;   99US-0142310.
PR   02-JUL-1999;   99US-0142311.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Williams LT,  Escobedo J,  Innis MA,  Garcia PD,  Klinger J,  Kassam A;
PI   Reinhard C,  Randazzo F,  Kennedy GC,  Pot D,  Lamson G,  Drmanac R;
PI   Crkenjakov R,  Drmanac S,  Dickson M,  Labat I,  Leshkowitiz D;
PI   Kita D,  Garcia V,  Jones LW,  Strache-Crain B;
XX
DR   WPI; 2001-091805/10.
XX
PT   Library of polynucleotides for diagnosing a cancerous state of a
PT   mammalian cell and detecting cancer, particularly of the colon or
PT   prostate, comprises 3351 human polynucleotide sequences -
XX
PS   Claim 9; Page 796; 1046pp; English.
XX
CC   The present sequence is one of 3351 sequences in a library of human
CC   polynucleotides. The library is used to detect differentially expressed
CC   genes correlated with a cancerous state of a mammalian cell and can
CC   detect colon, prostate, breast and lung cancer. The library can be used
CC   to produce probes for detection of mRNA and to produce additional copies
CC   of the polynucleotides. The probes can be used for chromosome mapping of
CC   the polynucleotide and for detection of transcription levels. Ribozymes
CC   or antisense oligonucleotides can be generated. The polynucleotides and
CC   their gene products are used as genetic or biochemical markers (e.g. in
CC   blood or tissues) that will detect the earliest changes along the
CC   carcinogenesis pathway and/or monitor the efficacy of therapies and
CC   preventive interventions. The polynucleotides, polypeptides and
CC   antibodies against them can be used in pharmaceutical compositions to
CC   treat the cancers and proliferative disorders such as neoplasia,
CC   dysplasia and hyperplasia.
XX
SQ   Sequence 373 BP; 127 A; 84 C; 90 G; 72 T; 0 other;

  Query Match             22.1%;  Score 291.2;  DB 22;  Length 373;
  Best Local Similarity   91.7%;  Pred. No. 1.2e-67;
  Matches  308;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy        172 TGAACGAAACAAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATC 231
              || |  | | ||  | ||| || |   |       |   | | ||||||||||||||||||
Db         37 TGGAGCAGAGAAAGAGGAATCATAGAGGTGCCAAAGGAACACAGACATAATGATGTCATC 96

Qy        232 CAAGCCAACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAG 291
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         97 CAAGCCAACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAG 156

Qy        292 CTTTATGGCTCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCA 351
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        157 CTTTATGGCTCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCA 216

Qy        352 GGGTGCTCAGCGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCA 411
              ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        217 GGGTGCTCAGCGTGCTCAGCCCTATGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCA 276

Qy        412 ACCGGGTCAAGGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTT 471
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        277 ACCGGGTCAAGGAAATATATAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTT 336

Qy        472 TAAAGAAGAAGCAAAGGCACTAGGGGTGATCCAGAT 507
              |||||||||||| ||||||||||||||||||||||
Db        337 TAAAGAAGAAGCTTAGGCACTAGGGGTGATCCAGAT 372


RESULT 10
AAS37423
ID   AAS37423 standard; cDNA; 383 BP.
XX
AC   AAS37423;
```

XX
DT 17-DEC-2001 (first entry)
XX
DE Novel human diagnostic and therapeutic gene #481.
XX
KW Human; cancer; breast; lung; colon; prostate; cytostatic; diagnostic; ss.
XX
OS Homo sapiens.
XX
PN WO200166753-A2.
XX
PD 13-SEP-2001.
XX
PF 09-MAR-2001; 2001WO-US07787.
XX
PR 09-MAR-2000; 2000US-0188609.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Kassam A, Lamson G;
PI Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones WL, Stache-Crain B;
XX
DR WPI; 2001-530177/58.
XX
PT New polynucleotides and polypeptides, useful for diagnosis and
PT treatment of breast, lung and colon cancer -
XX
PS Claim 1; Page 717; 1193pp; English.
XX
CC The invention relates to new polynucleotides and polypeptides, useful for
CC diagnosis and treatment of breast, lung and colon cancer. The sequences
CC can be used in detecting differentially expressed genes correlated with a
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample derived from a
CC cell suspected of being cancerous. They can also be used to inhibit
CC tumour growth by modulating expression of a gene product. AAS36943-
CC AAS39338 represent novel human diagnostic and therapeutic coding
CC sequences of the invention.
XX
SQ Sequence 383 BP; 129 A; 82 C; 89 G; 82 T; 1 other;

Query Match 21.9%; Score 288.6; DB 22; Length 383;
Best Local Similarity 90.1%; Pred. No. 6.2e-67;
Matches 309; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

```
Qy        172 TGAACGAAACAAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATC 231
              || |  | | ||  ||||| || |   |      |   |  ||||||||||||||| |
Db         41 TGGAGCACAGAAAGAAGAATCATAGAGGTGCCAAAGGAACTTAGACATAATGATGTCGTT 100

Qy        232 CAAGCCAACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAG 291
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        101 CAAGCCAACAAGCCAAGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAG 160

Qy        292 CTTTATGGCTCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCA 351
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        161 TTTTATGGCTCCTGGATTTCAACAGGCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCA 220

Qy        352 GGGTGCTCAGCGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCA 411
              ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        221 TGGTGCTCAGCGTGCTCAGCCCTATGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCA 280

Qy        412 ACCGGGTCAAGGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTT 471
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        281 ACCGGGTCAAGGAAATATATAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTT 340

Qy        472 TAAAGAAGAAGCAAAGGCACTAGGGGTGATCCAGATCATGGTT 514
              ||||||||||||||||||||| ||||||||| ||||||||
Db        341 TAAAGAAGAAGCAAAGGCACTATGGGTGATCCACATCATGGTT 383
```

RESULT 11
AAF65997
ID AAF65997 standard; cDNA; 370 BP.
XX
AC AAF65997;
XX
DT 09-APR-2001 (first entry)
XX
DE Novel human polynucleotide, SEQ ID NO: 1753.
XX
KW Human; cytostatic; gene therapy; colon cancer; prostate cancer;
KW breast cancer; lung cancer; cancer detection; ss.
XX
OS Homo sapiens.
XX
PN WO200102568-A2.
XX
PD 11-JAN-2001.
XX
PF 30-JUN-2000; 2000WO-US18374.
XX
PR 02-JUL-1999; 99US-0142310.
PR 02-JUL-1999; 99US-0142311.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;
PI Crkenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitiz D;
PI Kita D, Garcia V, Jones LW, Strache-Crain B;
XX
DR WPI; 2001-091805/10.
XX
PT Library of polynucleotides for diagnosing a cancerous state of a
PT mammalian cell and detecting cancer, particularly of the colon or
PT prostate, comprises 3351 human polynucleotide sequences -
XX
PS Claim 9; Page 795; 1046pp; English.
XX
CC The present sequence is one of 3351 sequences in a library of human
CC polynucleotides. The library is used to detect differentially expressed
CC genes correlated with a cancerous state of a mammalian cell and can
CC detect colon, prostate, breast and lung cancer. The library can be used
CC to produce probes for detection of mRNA and to produce additional copies
CC of the polynucleotides. The probes can be used for chromosome mapping of
CC the polynucleotide and for detection of transcription levels. Ribozymes
CC or antisense oligonucleotides can be generated. The polynucleotides and
CC their gene products are used as genetic or biochemical markers (e.g. in
CC blood or tissues) that will detect the earliest changes along the
CC carcinogenesis pathway and/or monitor the efficacy of therapies and
CC preventive interventions. The polynucleotides, polypeptides and
CC antibodies against them can be used in pharmaceutical compositions to
CC treat the cancers and proliferative disorders such as neoplasia,
CC dysplasia and hyperplasia.
XX
SQ Sequence 370 BP; 131 A; 83 C; 89 G; 67 T; 0 other;

Query Match 21.0%; Score 276; DB 22; Length 370;
Best Local Similarity 92.1%; Pred. No. 1.4e-63;
Matches 291; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

```
Qy        172 TGAACGAAACAAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATC 231
              || .|  | | ||  | |||||| |   |      |   | ||||||||||||||||||
Db         37 TGGAGCAGAGAAAGAGGAAACATAGAGGTGCCAAAGGAACAAAGACATAATGATGTCATC 96

Qy        232 CAAGCCAACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAG 291
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         97 CAAGCCAACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAG 156

Qy        292 CTTTATGGCTCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCA 351
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        157 CTTTATGGCTCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCA 216
```

```
Qy        352 GGGTGCTCAGCGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCA 411
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db        217 GGGTGCTCAGCGTGCTCAGCCCTATGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCA 276

Qy        412 ACCGGGTCAAGGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTT 471
              |||||| |||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        277 ACCGGGGCAAGGAAATATATAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTT 336

Qy        472 TAAAGAAGAAGCAAAG 487
              ||||||||||||||||
Db        337 TAAAGAAGAAGCAAAG 352
```

RESULT 12
AAV48114
ID AAV48114 standard; cDNA; 1236 BP.
XX
AC AAV48114;
XX
DT 27-OCT-1998 (first entry)
XX
DE Nucleotide sequence encoding clone HTPEF86 of TM4SF superfamily.
XX
KW ds; human; receptor; immune disorder; cancers; blood disorder;
KW juvenile rheumatoid arthritis; Graves disease.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 73..825
FT /*tag= a
FT /product= "Clone HTPEF86"
XX
PN WO9831799-A2.
XX
PD 23-JUL-1998.
XX
PF 21-JAN-1998; 98WO-US00959.
XX
PR 21-JAN-1997; 97US-0034205.
PR 21-JAN-1997; 97US-0034204.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Gentz RL, Ni J, Rosen CA;
XX
DR WPI; 1998-427559/36.
DR P-PSDB; AAW61619.
XX
PT New isolated poly:nucleotide(s) and encoded receptor poly:peptide(s)
PT - used to develop products for diagnosing or treating e.g. immune
PT disorders, cancers, blood disorders or immuno-compromised disease
PT states
XX
PS Claim 2; Page 38-39; 79pp; English.
XX
CC Clone HTPEF86 is a member of the TM4SF receptor superfamily. The
CC products generated using the receptor can be used for treating abnormal
CC conditions related to both an excess of and insufficient amounts of
CC receptor activity. They can be used in the treatment of e.g. immune
CC disorders, cancers, blood disorders, juvenile rheumatoid arthritis,
CC Graves disease or immunocompromised disease states. The products can
CC also be used for detection and diagnosis.
XX
SQ Sequence 1236 BP; 315 A; 320 C; 276 G; 325 T; 0 other;

Query Match 6.2%; Score 81.2;. DB 19; Length 1236;
Best Local Similarity 58.3%; Pred. No. 2.3e-11;
Matches 190; Conservative 0; Mismatches 118; Indels 18; Gaps 2;

```
Qy        590 TCTACTGCTGTTATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGC 649
              |||| | |  |    || || |  || |||||||| ||| | |  ||||| || || ||

Db        370 TCTATTTCATTCTACGGAGGCTTTCCCTTCTGGGGAGGCTTGTGGTTTATCATTTCAGGA 429

Qy        650 TCTCTCTCTGTGTCAGCATCCAAGGAGCT---TTCCCGTTGTCTGGTGAAAGGCAGCCTG 706
              |||||||| ||| |||||   ||  |||    |||   ||| ||| ||   |||||  ||
Db        430 TCTCTCTCCGTGGCAGCAGAAAATCAGCCATATTCTTATTGCCTGCTGTCTGGCAGTTTG 489

Qy        707 GGAATGAACATTGTTAGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGAT 766
              ||  ||||||| || ||| | ||||   |     ||||||| || ||  |  |    |||
Db        490 GGCTTGAACATCGTCAGTGCAATCTGCTCTGCAGTTGGAGTCATACTCTTCATCACAGAT 549

Qy        767 ATGTGCAT--------------CAATGGGGTAGCTGGCCAAGACTACTGGGCCGTGCTT 811
               |  | ||              | |    |    |   |   ||  |||||  |||
Db        550 CTAAGTATTCCCCACCCATATGCCTACCCCGACTATTATCCTTACGCCTGGGGTGTGAAC 609

Qy        812 TCTGGAAAAGGCATTTCAGCCACGCTGATGATCTTCTCCCTCTTGGAGTTCTTCGTAGCT 871
              ||||||  |  ||||| | |  |||| || |||||| |||| |||||||   | | ||
Db        610 CCTGGAATGGCGATTTCTGGCGTGCTGCTGGTCTTCTGCCTCCTGGAGTTTGGCATCGCA 669

Qy        872 TGTGCCACAGCCCATTTTGCCAACCA 897
              || ||  |  |||| |||| |  |||
Db        670 TGCGCATCTTCCCACTTTGGCTGCCA 695
```

RESULT 13
AAZ51573
ID AAZ51573 standard; cDNA; 1336 BP.
XX
AC AAZ51573;
XX
DT 21-JUN-2000 (first entry)
XX
DE Human cell surface immunomodulator-2 (CSIMM-2) encoding cDNA.
XX
KW Cell surface immunomodulator-2; CSIMM-2; cell proliferation; CD20;
KW differentiation; signal transduction; drug screening; prevention;
KW treatment; cancer; leukaemia; melanoma; immune disorder; AIDS;
KW rheumatoid arthritis; asthma; atherosclerosis; diabetes mellitus;
KW emphysema; irritable bowel syndrome; multiple sclerosis; diagnosis;
KW osteoporosis; psoriasis; microbial infections; cytostatic; antiarthritic;
KW antiasthmatic; immunosuppressive; antiarteriosclerotic; anti-HIV;
KW antidiabetic; antiinflammatory; neuroprotective; osteopathic;
KW antipsoriatic; antimicrobial; human; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 174..926
FT /*tag= a
FT /product= "Cell surface immunomodulator-2"
FT sig_peptide 174..239
FT /*tag= b
FT mat_peptide 240..923
FT /*tag= c
FT /product= "Mature cell surface immunomodulator-2"
FT misc_binding 801..830
FT /*tag= d
FT /bound_moiety= "Hybridisation probe"
XX
PN WO200011150-A1.
XX
PD 02-MAR-2000.
XX
PF 24-AUG-1999; 99WO-US19386.
XX
PR 25-AUG-1998; 98US-0139687.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Lal P, Corley NC, Gorgone GA, Guegler KJ, Patterson C, Baughn MR;
XX
DR WPI; 2000-246561/21.
DR P-PSDB; AAY70432.
XX

PT New human cell surface immunomodulatory polypeptides and
PT polynucleotides useful for diagnosis, prevention and treatment of
PT cancer and immune disorders -
XX
PS Claim 9; Page 63-64; 70pp; English.
XX
CC The present sequence is a cDNA identified in Incyte clone 2721336,
CC derived from lung tumour cDNA library (LUNGTUT10). It encodes
CC human cell surface immunomodulator-2 (CSIMM-2), which is a
CC regulator of cell proliferation, differentiation, cell-cell
CC communication and signal transduction. It shows homology to cell surface
CC antigen, CD20. CSIMM can be used for drug screening, prevention and
CC treatment of cancers such as leukaemia and melanoma, immune disorders
CC such as AIDS, rheumatoid arthritis, asthma, atherosclerosis, diabetes
CC mellitus, emphysema, irritable bowel syndrome, multiple sclerosis,
CC osteoporosis, psoriasis and microbial infections. CSIMM polynucleotide
CC may be used for diagnosis of CSIMM-associated diseases and as source of
CC probes useful in mapping naturally occurring genomic sequences.
XX
SQ Sequence 1336 BP; 336 A; 354 C; 296 G; 350 T; 0 other;

Query Match 6.2%; Score 81.2; DB 21; Length 1336;
Best Local Similarity 58.3%; Pred. No. 2.3e-11;
Matches 190; Conservative 0; Mismatches 118; Indels 18; Gaps 2;

```
Qy      590 TCTACTGCTGTTATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGC 649
            |||| | |  |    || || |  || |||||||| ||| | |  ||||| || || ||
Db      471 TCTATTTCATTCTACGGAGGCTTTCCCTTCTGGGGAGGCTTGTGGTTTATCATTTCAGGA 530

Qy      650 TCTCTCTCTGTGTCAGCATCCAAGGAGCT---TTCCCGTTGTCTGGTGAAAGGCAGCCTG 706
            |||||||| ||| |||||   ||  |||    |||   ||| ||| ||   ||||| ||
Db      531 TCTCTCTCCGTGGCAGCAGAAAATCAGCCATATTCTTATTGCCTGCTGTCTGGCAGTTTG 590

Qy      707 GGAATGAACATTGTTAGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGAT 766
            ||  ||||||| || ||| | ||||   |     ||||||| || ||  |  |    |||
Db      591 GGCTTGAACATCGTCAGTGCAATCTGCTCTGCAGTTGGAGTCATACTCTTCATCACAGAT 650

Qy      767 ATGTGCAT--------------CAATGGGGTAGCTGGCCAAGACTACTGGGCCGTGCTT 811
             |  | ||               | |    |    |   |   ||  |||||  |||
Db      651 CTAAGTATTCCCCACCCATATGCCTACCCCGACTATTATCCTTACGCCTGGGGTGTGAAC 710

Qy      812 TCTGGAAAAGGCATTTCAGCCACGCTGATGATCTTCTCCCTCTTGGAGTTCTTCGTAGCT 871
             ||||||  |  ||||| | |  |||| || |||||| |||| |||||||   | | ||
Db      711 CCTGGAATGGCGATTTCTGGCGTGCTGCTGGTCTTCTGCCTCCTGGAGTTTGGCATCGCA 770

Qy      872 TGTGCCACAGCCCATTTTGCCAACCA 897
            || ||  |  |||| |||| |  |||
Db      771 TGCGCATCTTCCCACTTTGGCTGCCA 796
```

RESULT 14
ABK39631
ID ABK39631 standard; cDNA; 1336 BP.
XX
AC ABK39631;
XX
DT 21-MAY-2002 (first entry)
XX
DE cDNA encoding L985P lung tumour associated protein.
XX
KW Lung tumour; cancer; T cell; immune response stimulator;
KW cytostatic; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200204514-A2.
XX
PD 17-JAN-2002.
XX
PF 10-JUL-2001; 2001WO-US22058.
XX
PR 11-JUL-2000; 2000US-0614124.
PR 29-AUG-2000; 2000US-0651563.
PR 08-SEP-2000; 2000US-0658824.
PR 26-SEP-2000; 2000US-0671325.
PR 06-OCT-2000; 2000US-0677419.
PR 30-OCT-2000; 2000US-0702705.
PR 13-DEC-2000; 2000US-0736457.
PR 03-MAY-2001; 2001US-0849626.
XX
PA (CORI-) CORIXA CORP.
XX
PI Wang T, Watanabe Y, Henderson RA, Johnson JC, Retter MW;
PI Marnerakis M, Carter D, Fanger GR, Vedvick TS, Bangur CS;
PI Mcnabb A, Wang A, Fanger N, Switzer A, Mcneill PD, Clapper JD;
XX
DR WPI; 2002-164634/21.
DR P-PSDB; AAU85539.
XX
PT Novel polynucleotide encoding a lung tumour polypeptide useful for
PT stimulating and/or expanding T cells specific for a tumour protein -
XX
PS Claim 7; SEQ ID No 1676; 223pp; English.
XX
CC The invention describes an isolated polynucleotide and polypeptide
CC useful for stimulating and/or expanding T cells specific for a tumour
CC protein for determining the presence of a cancer in a patient. A
CC composition containing the polynucleotide and/or polypeptide is useful
CC for treating a lung cancer in a patient. The polypeptide is useful for
CC removing tumour cells from a biological sample. The polynucleotide is
CC also useful as probe or primer to detect the level of mRNA encoding a
CC tumour protein. This sequence encodes a lung tumour associated protein
CC or protein fragment, described in the method of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1336 BP; 336 A; 354 C; 296 G; 350 T; 0 other;

Query Match 6.2%; Score 81.2; DB 24; Length 1336;
Best Local Similarity 58.3%; Pred. No. 2.3e-11;
Matches 190; Conservative 0; Mismatches 118; Indels 18; Gaps 2;

```
Qy      590 TCTACTGCTGTTATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGC 649
            |||| | |  |    || || |  || |||||||| ||| | |  ||||| || || ||
Db      471 TCTATTTCATTCTACGGAGGCTTTCCCTTCTGGGGAGGCTTGTGGTTTATCATTTCAGGA 530

Qy      650 TCTCTCTCTGTGTCAGCATCCAAGGAGCT---TTCCCGTTGTCTGGTGAAAGGCAGCCTG 706
            |||||||| ||| |||||   ||  |||    |||   ||| ||| ||   ||||| ||
Db      531 TCTCTCTCCGTGGCAGCAGAAAATCAGCCATATTCTTATTGCCTGCTGTCTGGCAGTTTG 590

Qy      707 GGAATGAACATTGTTAGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGAT 766
            ||  ||||||| || ||| | ||||   |     ||||||| || ||  |  |    |||
Db      591 GGCTTGAACATCGTCAGTGCAATCTGCTCTGCAGTTGGAGTCATACTCTTCATCACAGAT 650

Qy      767 ATGTGCAT--------------CAATGGGGTAGCTGGCCAAGACTACTGGGCCGTGCTT 811
             |  | ||               | |    |    |   |   ||  |||||  |||
Db      651 CTAAGTATTCCCCACCCATATGCCTACCCCGACTATTATCCTTACGCCTGGGGTGTGAAC 710

Qy      812 TCTGGAAAAGGCATTTCAGCCACGCTGATGATCTTCTCCCTCTTGGAGTTCTTCGTAGCT 871
             ||||||  |  ||||| | |  |||| || |||||| |||| |||||||   | | ||
Db      711 CCTGGAATGGCGATTTCTGGCGTGCTGCTGGTCTTCTGCCTCCTGGAGTTTGGCATCGCA 770

Qy      872 TGTGCCACAGCCCATTTTGCCAACCA 897
            || ||  |  |||| |||| |  |||
Db      771 TGCGCATCTTCCCACTTTGGCTGCCA 796
```

RESULT 15
AAC99077
ID AAC99077 standard; cDNA; 1374 BP.
XX
AC AAC99077;
XX
DT 09-MAR-2001 (first entry)

```
XX
DE   Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:305.
XX
KW   Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
KW   detection; diagnosis; identification; cytostatic; neuroprotective;
KW   nootropic; immunomodulatory; relaxant; contraceptive; gynaecological;
KW   antiinflammatory; cardiant; gene therapy; chromosome mapping;
KW   linkage analysis; tissue identification; tissue typing; forensic;
KW   neural; immune system; muscular; reproductive; gastrointestinal;
KW   pulmonary; cardiovascular; renal; proliferative; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055320-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05989.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-579444/54.
DR   P-PSDB; AAB54312.
XX
PT   New nucleic acid that is a pancreatic cancer antigen for preventing,
PT   treating, or ameliorating a medical condition, particular pancreatic
PT   cancer, or for use in assays for diagnosing a pathological condition -
XX
PS   Claim 1; Page 731; 1379pp; English.
XX
CC   AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC   proteins, called pancreatic cancer antigens, given in AAB54008 to
CC   AAB54466. The human pancreatic cancer antigens have cytostatic,
CC   neuroprotective, nootropic, immunomodulatory, relaxant, contraceptive,
CC   gynaecological, cardiant and antiinflammatory activities, and can be used
CC   in gene therapy. The polynucleotide and proteins can be used for
CC   preventing, treating, or ameliorating a medical condition or in assays
CC   for diagnosing a pathological condition or a susceptibility to one in a
CC   subject. Binding partners to the proteins and the activity of the
CC   proteins can be identified. The pancreatic cancer antigens can be used to
CC   detect, treat or prevent pancreatic disorders, especially cancer.
CC   Agonists and antagonists to the antigens can be screened for. The
CC   pancreatic cancer antigen polynucleotides can be used to design nucleic
CC   acid hybridisation probes that can be used in chromosome mapping, linkage
CC   analysis, tissue identification and/or typing and a variety of forensic
CC   and diagnostic methods. The proteins can be used to generate antibodies
CC   which are used to purify, detect and target the polypeptides, including
CC   both in vivo and in vitro diagnostic and therapeutic methods. The
CC   proteins can be used to treat or prevent neural, immune system, muscular,
CC   reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC   proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence 1374 BP; 361 A; 358 C; 301 G; 353 T; 1 other;

  Query Match             6.2%;  Score 81.2;  DB 21;  Length 1374;
  Best Local Similarity   58.3%;  Pred. No. 2.4e-11;
  Matches  190;  Conservative    0;  Mismatches  118;  Indels   18;  Gaps    2;

Qy        590 TCTACTGCTGTTATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGC 649
              |||| | |  |    || || |  || ||||||| ||| | |  ||||| || || ||
Db        501 TCTATTTCATTCTACGGAGGCTTTCCCTTCTGGGGAGGCTTGTGGTTTATCATTTCAGGA 560

Qy        650 TCTCTCTCTGTGTCAGCATCCAAGGAGCT---TTCCCGTTGTCTGGTGAAAGGCAGCCTG 706
              |||||||| ||| |||||   ||  |||    |||   ||| ||| ||   |||||  ||
Db        561 TCTCTCTCCGTGGCAGCAGAAAATCAGCCATATTCTTATTGCCTGCTGTCTGGCAGTTTG 620

Qy        707 GGAATGAACATTGTTAGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGAT 766
              ||  ||||||| || ||| | ||||   |     ||||||| || ||  |  |    |||

Db        621 GGCTTGAACATCGTCAGTGCAATCTGCTCTGCAGTTGGAGTCATACTCTTCATCACAGAT 680

Qy        767 ATGTGCAT----------------CAATGGGGTAGCTGGCCAAGACTACTGGGCCGTGCTT 811
              |  | ||                | |    |    |    ||  |||||  |||
Db        681 CTAAGTATTCCCCACCCATATGCCTACCCCGACTATTATCCTTACGCCTGGGGTGTGAAC 740

Qy        812 TCTGGAAAAGGCATTTCAGCCACGCTGATGATCTTCTCCCTCTTGGAGTTCTTCGTAGCT 871
              ||||||  |  ||||| | |  |||| || |||||| |||| |||||||   | | ||
Db        741 CCTGGAATGGCGATTTCTGGCGTGCTGCTGGTCTTCTGCCTCCTGGAGTTTGGCATCGCA 800

Qy        872 TGTGCCACAGCCCATTTTGCCAACCA 897
              || ||  |  |||| |||| |  |||
Db        801 TGCGCATCTTCCCACTTTGGCTGCCA 826


Search completed: July 16, 2003, 11:09:57
Job time : 227.967 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 16, 2003, 08:51:42 ; Search time 1437.76 Seconds
(without alignments)
14835.236 Million cell updates/sec

Title: US-09-981-353-81
Perfect score: 1317
Sequence: 1 aaggaacaaagtaagtccat.....tttttccttaataaaatgtc 1317

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	728.2	55.3	859	13	BI760204	BI760204 603045133
c	2	369.2	28.0	417	10	AW842562	AW842562 MR2-CN003
c	3	368.2	28.0	415	10	AW842606	AW842606 MR2-CN003
	4	200.6	15.2	452	14	F15017	F15017 SSO4B08 Por
	5	132.2	10.0	418	17	AQ213250	AQ213250 HS_3249_B
	6	96.4	7.3	176	13	BG983751	BG983751 MR4-CN014
	7	87.4	6.6	185	13	BG981760	BG981760 MR3-CN014
	8	79.6	6.0	880	13	BI766299	BI766299 603052775
	9	77.6	5.9	687	13	BM488591	BM488591 pgm2n.pk0
	10	75.6	5.7	673	10	BB629370	BB629370 BB629370
	11	74.8	5.7	515	12	BG434337	BG434337 602506331
	12	74.8	5.7	594	10	AV717594	AV717594 AV717594
	13	74.8	5.7	619	10	AV661743	AV661743 AV661743
	14	74.8	5.7	680	12	BG546745	BG546745 602574115
	15	74.8	5.7	735	10	AV715678	AV715678 AV715678
	16	74.8	5.7	869	13	BI766801	BI766801 603056712
	17	74.8	5.7	876	13	BI771044	BI771044 603055393
	18	74.8	5.7	934	13	BI821317	BI821317 603038160
	19	74.4	5.6	597	10	AV716310	AV716310 AV716310
	20	74.4	5.6	860	9	AL551751	AL551751 AL551751
	21	73.2	5.6	656	14	BM781186	BM781186 MLN1_5_B1
	22	71.2	5.4	481	9	AA160278	AA160278 zo87b11.r
	23	70.6	5.4	584	13	BI824549	BI824549 603035154
	24	70.6	5.4	1017	11	AK008099	AK008099 Mus muscu
	25	69.6	5.3	853	12	BF574865	BF574865 602132285
	26	69.6	5.3	885	13	BI911270	BI911270 603062843
	27	69.2	5.3	619	17	BH338441	BH338441 CH230-74J
	28	69	5.2	436	9	AI635515	AI635515 ts95b02.x
	29	69	5.2	540	9	AI552228	AI552228 vf70b04.y
	30	68.4	5.2	577	10	AV717053	AV717053 AV717053
c	31	67.8	5.1	773	9	AI263085	AI263085 qz35a05.x
	32	67.2	5.1	608	12	BF042446	BF042446 BP250003A
	33	67	5.1	528	12	BF868999	BF868999 QV3-ET010
c	34	67	5.1	648	9	AI741604	AI741604 wg28d01.x
	35	66.4	5.0	449	9	AA469742	AA469742 vf70b04.r
	36	66.4	5.0	582	10	AV714513	AV714513 AV714513
	37	65	4.9	504	12	BF522968	BF522968 UI-R-C2p-
	38	64.8	4.9	517	12	BG664654	BG664654 DRABGG07
	39	64.6	4.9	886	13	BI908327	BI908327 603066152
	40	63.6	4.8	717	13	BI771406	BI771406 603059433
	41	63.4	4.8	483	13	BM430393	BM430393 1Duo33F9.
	42	63.4	4.8	604	12	BF557830	BF557830 UI-R-C0-g
	43	63.2	4.8	491	12	BF556579	BF556579 UI-R-E1-f
	44	63.2	4.8	507	9	AI792758	AI792758 ol01d12.y
	45	63.2	4.8	686	10	AV733821	AV733821 AV733821

ALIGNMENTS

RESULT 1
BI760204
LOCUS BI760204 859 bp mRNA linear EST 25-SEP-2001
DEFINITION 603045133F1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5185424 5', mRNA sequence.
ACCESSION BI760204
VERSION BI760204.1 GI:15751782
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 859)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11463 row: a column: 09
High quality sequence stop: 792.
FEATURES Location/Qualifiers
source 1. .859

```
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5185424"
                     /clone_lib="NIH_MGC_116"
                     /lab_host="DH10B"
                     /note="Organ: pooled colon, kidney, stomach; Vector:
                     pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
                     source anonymous pool of 3 colons, age 26 yo male, 49 yo
                     female, 71 yo male colon; 46 yo male kidney, and pool of 2
                     stomachs, 62 yo male and 70 yo female.  Library is
                     oligo-dT primed and  directionally cloned (EcoRV site is
                     destroyed upon cloning).  Average insert size 1.4 kb,
                     insert size range 1-3 kb.  Library is normalized and
                     enriched for full-length clones and was constructed by C.
                     Gruber  (Invitrogen).  Research Genetics tracking code
                     023. Note: this is a NIH_MGC Library."
BASE COUNT      229 a    198 c    192 g    240 t
ORIGIN

  Query Match             55.3%;  Score 728.2;  DB 13;  Length 859;
  Best Local Similarity   92.9%;  Pred. No. 1.2e-141;
  Matches  785;  Conservative    0;  Mismatches   58;  Indels    2;  Gaps    2;

Qy        178 AAACAAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATCCAAGCC 237
              | | ||  | ||||||  |    |       |   | ||||||||||||||||||||||||||
Db          3 AGAGAAAGAGGAAACATAGAGGTGCCAAAGGAACAAAGACATAATGATGTCATCCAAGCC 62

Qy        238 AACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAGCTTTAT 297
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         63 AACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAGCTTTAT 122

Qy        298 GGCTCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCAGGGTGC 357
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 GGCTCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCAGGGTGC 182

Qy        358 TCAGCGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCAACCGGG 417
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 TCAGCGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCAACCGGG 242

Qy        418 TCAAGGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTTTAAAGA 477
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 TCAAGGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTTTAAAGA 302

Qy        478 AGAAGCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGGTTTTGG 537
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        303 AGAAGCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGGTTTCGG 362

Qy        538 AATTGTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCTACTGC 597
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 AATTGTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCTACTGC 422

Qy        598 TGTTATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCTCTCTC 657
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 TGTTATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCTCTCTC 482

Qy        658 TGTGTCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATGAACAT 717
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        483 TGTGTCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATGAACAT 542

Qy        718 TGTTAGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGCATCAA 777
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        543 TGTTAGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGCATCAA 602

Qy        778 TGGGGTAGCTGGCCAAGACTACTGGGCCGTGCTTTCTGGAAAAGGCA-TTTCAGCCACGC 836
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        603 TGGGGTAGCTGGCCAAGACTACTGGGCCGTGCTTTCTGGAAAAGGCATTTTCAGCCACGC 662

Qy        837 TGATGATCTTCTCCCTCTT-GGAGTTCTTCGTAGCTTGTGCCACAGCCCATTTTGCCAAC 895
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        663 TGATGATCTTCTCCCTCTTGGGAGTTCTTCGTAGCTTGTGCCACAGCCCATTTTGCCAAC 722

Qy        896 CAAGCAAACACCACAACCAATATGTCTGTCCTGGTTATTCCAAATATGTATGAAAGCAAC 955
              ||||||||  |   |    || | | |     | |  | | ||   |||||||||| ||
Db        723 CAAGCAAAACCACAATAATATGTCTGTCCTGGGTTATTCCAAATATTGTATGAAAGCCAC 782

Qy        956 CCTGTGACACCAGCGTCTTCTTCAGCTCCTCCCAGATGCAACAACTACTCAGCTAATGCC 1015
              ||||||||||||||||||  |    ||||||||||||||||   | ||||||||||||
Db        783 CCTGTGACACCAGCGTCTTTCTTCAGTCCTCCCAGATGCAACACTAATTCAGCTAATGCC 842

Qy       1016 CCTAA 1020
                |  ||
Db         843 CTAAA 847


RESULT 2
AW842562/c
LOCUS       AW842562                417 bp    mRNA    linear   EST 18-MAY-2000
DEFINITION  MR2-CN0035-170300-201-a10 CN0035 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW842562
VERSION     AW842562.1  GI:7936545
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 417)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=MR2-CN0035-170
            300-201-a10&t3=2000-03-17&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 382.
FEATURES             Location/Qualifiers
     source          1. .417
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="CN0035"
                     /dev_stage="Adult"
                     /note="Organ: colon_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      123 a     94 c    110 g     90 t
ORIGIN

  Query Match             28.0%;  Score 369.2;  DB 10;  Length 417;
  Best Local Similarity   97.7%;  Pred. No. 8.1e-67;
  Matches  385;  Conservative    0;  Mismatches    8;  Indels    1;  Gaps    1;

Qy        607 TGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCTCTCTCTGTGTCAGC 666
              |||||||  |||||||||||||||  |||||  | ||||||||||||||||||||||||||
Db        416 TGGATACTCATTCTGGGGTGGCCATTCTTACAATATCTCTGGCTCTCTCTCTGTGTCAGC 357

Qy        667 ATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATGAACATTGTTAGTTC 726
              |||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        356 ATCCAAGGAGCCTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATGAACATTGTTAGTTC 297

Qy        727 TATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGCATCAATGGGGTAGC 786
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        296 TATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGCATCCATGGGGTAGC 237

Qy        787 TGGCCAAGACTACTGGGCCGTGCTTTCTGGAAAAGGCATTTCAGCCACGCTGATGATCTT 846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        236 TGGCCAAGACTACTGGGCCGTGCTTTCTGGAAAAGGCATTTCAGCCACGCTGATGATCTT 177

Qy        847 CTCCCTCTTGGAGTTCTTCGTAGCTTGTGCCACAGCCCATTTTGCCAA-CCAAGCAAACA 905
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        176 CTCCCTCTTGGAGTTCTTCGTAGCTTGTGCCACAGCCCATTTTGCCAATCCAAGCAAACA 117

Qy        906 CCACAACCAATATGTCTGTCCTGGTTATTCCAAATATGTATGAAAGCAACCCTGTGACAC 965
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        116 CCACAACCCATATGTCTGTCCTGGTTATTCCAAATATGTATGAAAGCAACCCTGTGACAC 57

Qy        966 CAGCGTCTTCTTCAGCTCCTCCCAGATGCAACAA 999
              |||||||||||||||||||||||||||||||||
Db         56 CAGCGTCTTCTTCAGCTCCTCCCAGATGCAACAA 23
```

RESULT 3
AW842606/c

LOCUS	AW842606 415 bp mRNA linear EST 18-MAY-2000
DEFINITION	MR2-CN0035-200300-203-h08 CN0035 Homo sapiens cDNA, mRNA sequence.
ACCESSION	AW842606
VERSION	AW842606.1 GI:7936589
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 415)
AUTHORS	Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE	20202663
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=MR2-CN0035-200300-203-h08&t3=2000-03-20&t4=1) Seq primer: puc 18 forward High quality sequence start: 53 High quality sequence stop: 415.

FEATURES	Location/Qualifiers
source	1. .415 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="CN0035" /dev_stage="Adult" /note="Organ: colon_normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

```
BASE COUNT      124 a     93 c    112 g     86 t
ORIGIN

  Query Match           28.0%;  Score 368.2;  DB 10;  Length 415;
  Best Local Similarity 97.7%;  Pred. No. 1.3e-66;
  Matches  384;  Conservative    0;  Mismatches    8;  Indels    1;  Gaps    1;

Qy        609 GATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCTCTCTCTGTGTCAGCAT 668
              |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db        415 GATACCCATTCT-GGGTGGCCTTTCTTTTATTATCTCTGGCTCTCTCTCTGTGTCAGCAT 357

Qy        669 CCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATGAACATTGTTAGTTCTA 728
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        356 CCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATGAACATTGTTAGTTCTA 297

Qy        729 TCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGCATCAATGGGGTAGCTG 788
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        296 TCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGCATCCATGGGGTAGCTG 237

Qy        789 GCCAAGACTACTGGGCCGTGCTTTCTGGAAAAGGCATTTCAGCCACGCTGATGATCTTCT 848
              || |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        236 TCCCAGACTACTGGGCCGTGCGTTCTGGAAAAGGCATTTCAGCCACGCTGATGATCTTCT 177

Qy        849 CCCTCTTGGAGTTCTTCGTAGCTTGTGCCACAGCCCATTTTGCCAACCAAGCAAACACCA 908
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        176 CCCTCTTGGAGTTCTTCGTAGCTTGTGCCACAGCCCATTTTGCCAACCAAGCATACACCA 117

Qy        909 CAACCAATATGTCTGTCCTGGTTATTCCAAATATGTATGAAAGCAACCCTGTGACACCAG 968
              |||||||||||||||||||||||||| |||||||||||||||||| ||||||||||||
Db        116 CAACCAATATGTCTGTCCTGGTTATTCCCAATATGTATGAAAGCAAGCCTGTGACACCAG 57

Qy        969 CGTCTTCTTCAGCTCCTCCCAGATGCAACAACT 1001
              |||||||||||||||||||||||||||||||| |
Db         56 CGTCTTCTTCAGCTCCTCCCAGATGCAACAAAT 24
```

RESULT 4
F15017

LOCUS	F15017 452 bp mRNA linear EST 09-SEP-1996
DEFINITION	SSO4B08 Porcine small intestine cDNA library Sus scrofa cDNA clone o4b08, mRNA sequence.
ACCESSION	F15017
VERSION	F15017.1 GI:972439
KEYWORDS	EST.
SOURCE	pig.
ORGANISM	Sus scrofa Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE	1 (bases 1 to 452)
AUTHORS	Wintero,A.K., Fredholm,M. and Davies,W.
TITLE	Evaluation and characterization of a porcine small intestine cDNA library: analysis of 839 clones
JOURNAL	Mamm. Genome 7 (7), 509-517 (1996)
MEDLINE	96327607
COMMENT	Contact: A.K. Winteroe Department of Animal Science and Animal Health, Division of Animal Genetics, The Royal Veterinary and Agricultural University Bulowsvej 13, 1870 Frederiksberg C, Denmark.

FEATURES	Location/Qualifiers
source	1. .452 /organism="Sus scrofa" /db_xref="taxon:9823" /clone="o4b08" /clone_lib="Porcine small intestine cDNA library" /note="directionally cloned cDNA in XL1-blue MRF'"

```
BASE COUNT      149 a    103 c    100 g     97 t      3 others
ORIGIN

  Query Match           15.2%;  Score 200.6;  DB 14;  Length 452;
  Best Local Similarity 69.6%;  Pred. No. 1e-31;
  Matches  288;  Conservative    0;  Mismatches  117;  Indels    9;  Gaps    1;

Qy        173 GAACGAAACAAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATCC 232
```

```
Db         48 GAGAGACAGAGAGAGGAAGCCAAGAAGTGCCAAAGGANCAGAGACAAAATGATGTCATCC 107

Qy        233 AAGCCAACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAGC 292

Db        108 AAGCCAACAACCTACCCTGGAGTGTATGGAACCACACCTGATCTTTACCCACCAAGTAAC 167

Qy        293 TTTATGGCTCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCAG 352

Db        168 TTTATGGTGCCTGGGTNTCAACAACCCCCAGGGTTCATCAACCCACGAATCCAAGTTCAG 227

Qy        353 GGTGCTCAGCGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCAA 412

Db        228 AGTAGCCAGGCACCT---------TTCATCGTCTCTCCAGGAATCCCCAATAACAGTCAG 278

Qy        413 CCGGGTCAAGGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTTT 472

Db        279 CAGGTTCAAGGANATATACAAATGGTAAATCCAGGTACAGGAAAAGCAGCCACAAACTTT 338

Qy        473 AAAGAAGAAGCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGGT 532

Db        339 AAAGAAGAAGCAAAGACACTAGGGGCAATCCAGATCCTGATTGGTTCAATGCACATTGGC 398

Qy        533 TTTGGAATTGTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTT 586

Db        399 TTCGGAATTATTCTGGGCTTAATGGGCCGCACTTATATTCAAGTTTTGGGTTTT 452


RESULT 5
AQ213250
LOCUS       AQ213250                418 bp    DNA     linear   GSS 18-SEP-1998
DEFINITION  HS_3249_B1_A05_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3249 Col=9 Row=B, DNA sequence.
ACCESSION   AQ213250
VERSION     AQ213250.1  GI:3624451
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 418)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3249  row: B  column: 9
            Class: BAC ends
            High quality sequence stop: 418.
FEATURES             Location/Qualifiers
     source          1. .418
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=3249 Col=9 Row=B"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      106 a     87 c     87 g    135 t      3 others
ORIGIN

  Query Match             10.0%;  Score 132.2;  DB 17;  Length 418;
  Best Local Similarity   93.8%;  Pred. No. 1.9e-17;
  Matches  137;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy        496 GGTGATCCAGATCATGGTTGGATTGATGCACATTGGTTTTGGAATTGTTTTGTGTTTAAT 555

Db        249 GGTGATCCAGATCATGGTTGGATTGATGCACATTGGTTNTGGAATTGTTTTGTGTTTAAT 308

Qy        556 ATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCTACTGCTGTTATTGGTGGATACCC 615

Db        309 ATCCTTCTCTTTTAGAGAAGCATTAAAGTTTGCCTCTACTGCTGTTATTGGTGGATACCC 368

Qy        616 ATTCTGGGGTGGCCTTTCTTTTATTA 641

Db        369 ATTCTGGGGTGGCCTCTCTGTGAGTA 394


RESULT 6
BG983751
LOCUS       BG983751                176 bp    mRNA    linear   EST 12-JUN-2001
DEFINITION  MR4-CN0147-140301-010-d05 CN0147 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BG983751
VERSION     BG983751.1  GI:14386486
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 176)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR4&t2=MR4-CN0147-
            140301-010-d05&t3=2001-03-14&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 173.
FEATURES             Location/Qualifiers
     source          1. .176
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="CN0147"
                     /dev_stage="Adult"
                     /note="Organ: colon_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       97 a     20 c     21 g     38 t
ORIGIN

  Query Match             7.3%;  Score 96.4;  DB 13;  Length 176;
  Best Local Similarity   99.0%;  Pred. No. 6e-10;
  Matches   97;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1220 GGCAAAGGTGAAGGATCAGAGGACTGAAAAATGATTCTGCAACTCTCTTAAAGTTAGAAA 1279
```

```
Db          9 GGCACAGGTGAAGGATCAGAGGACTGAAAAATGATTCTGCAACTCTCTTAAAGTTAGAAA 68

Qy       1280 TGTTTCTGTTCATATTACTTTTTCCTTAATAAAATGTC 1317
              |||||||||||||||||||||||||||||||||||||
Db         69 TGTTTCTGTTCATATTACTTTTTCCTTAATAAAATGTC 106


RESULT 7
BG981760
LOCUS       BG981760                185 bp    mRNA    linear   EST 12-JUN-2001
DEFINITION  MR3-CN0147-090201-014-e06 CN0147 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BG981760
VERSION     BG981760.1  GI:14384495
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 185)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR3&t2=MR3-CN0147-
            090201-014-e06&t3=2001-02-09&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 184.
FEATURES             Location/Qualifiers
     source          1. .185
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="CN0147"
                     /dev_stage="Adult"
                     /note="Organ: colon_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       75 a     24 c     37 g     49 t
ORIGIN

  Query Match             6.6%;  Score 87.4;  DB 13;  Length 185;
  Best Local Similarity   98.9%;  Pred. No. 4.5e-08;
  Matches   88;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1229 GAAGGATCAGAGGACTGAAAAATGATTCTGCAACTCTCTTAAAGTTAGAAATGTTTCTGT 1288
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         20 GGAGGATCAGAGGACTGAAAAATGATTCTGCAACTCTCTTAAAGTTAGAAATGTTTCTGT 79

Qy       1289 TCATATTACTTTTTCCTTAATAAAATGTC 1317
              |||||||||||||||||||||||||||||
Db         80 TCATATTACTTTTTCCTTAATAAAATGTC 108


RESULT 8
BI766299
LOCUS       BI766299                880 bp    mRNA    linear   EST 25-SEP-2001
DEFINITION  603052775F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5202426 5',
            mRNA sequence.
ACCESSION   BI766299
VERSION     BI766299.1  GI:15757877
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 880)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11507  row: e  column: 19
            High quality sequence stop: 845.
FEATURES             Location/Qualifiers
     source          1. .880
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5202426"
                     /clone_lib="NIH_MGC_122"
                     /lab_host="DH10B"
                     /note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
                     Site_1: NotI; Site_2: EcoRV (destroyed); RNA source
                     anonymous pool of 24 week female lung, 16 week female
                     spleen, and 20-22 week male spleens.  Library is oligo-dT
                     primed and directionally cloned (EcoRV site is destroyed
                     upon  cloning).  Average insert size 1.4 kb, insert size
                     range 1-3 kb. Library is normalized and enriched for
                     full-length clones and was  constructed by C. Gruber
                     (Invitrogen).  Research Genetics tracking  code 026. Note:
                     this is a NIH_MGC Library."
BASE COUNT      208 a    239 c    209 g    223 t      1 others
ORIGIN

  Query Match             6.0%;  Score 79.6;  DB 13;  Length 880;
  Best Local Similarity   58.0%;  Pred. No. 1.5e-06;
  Matches  189;  Conservative    0;  Mismatches  119;  Indels   18;  Gaps    2;

Qy        590 TCTACTGCTGTTATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGC 649
              |||| | |  |    || || |  || |||||||| ||| | |  ||||| || || ||
Db        440 TCTATTTCATTCTACGGAGGCTTTCCCTTCTGGGGAGGCTTGTGGTTTATCATTTCAGGA 499

Qy        650 TCTCTCTCTGTGTCAGCATCCAAGGAGCT---TTCCCGTTGTCTGGTGAAAGGCAGCCTG 706
              |||||||| ||| |||||   ||  |||    |||   ||| ||| ||   |||||  ||
Db        500 TCTCTCTCCGTGGCAGCAGAAAATCAGCCATATTCTTATTGCCTGCTGTCTGGCAGTTTG 559

Qy        707 GGAATGAACATTGTTAGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGAT 766
              ||  ||||||| || ||| | ||||   |   . ||||||| || ||  |  |    |||
Db        560 GGCTTGAACATCGTCAGTGCAATCTGCTCTGCAGTTGGAGTCATACTCTTCATCACAGAT 619

Qy        767 ATGTGCATCAATGGGGTAGCTGGC--------------CAAGACTACTGGGCCGTGCTT 811
               |  | ||    |   | || |                |   ||  |||||  |||
Db        620 CTAAGTATTCCCCGACCATATGCCTACCCCGACTATTATCCTTACGCCTGGGGTGTGAAC 679

Qy        812 TCTGGAAAAGGCATTTCAGCCACGCTGATGATCTTCTCCCTCTTGGAGTTCTTCGTAGCT 871
              ||||||  |  ||||| | |  |||| || |||||| |||| |||||||   | | ||
Db        680 CCTGGAATGGCGATTTCTGGCGTGCTGCTGGTCTTCTGCCTCCTGGAGTTTGGCATCGCA 739

Qy        872 TGTGCCACAGCCCATTTTGCCAACCA 897
              || ||  |  | ||  ||| |  |||
Db        740 TGCGCATCTTCACAACTTGGCTGCCA 765
```

```
RESULT 9
BM488591
LOCUS       BM488591                687 bp    mRNA    linear   EST 07-FEB-2002
DEFINITION  pgm2n.pk008.c22 Normalized Chicken Breast Muscle, Leg Muscle, and
            Epiphyseal Growth Plate cDNA library (pgm2n) Gallus gallus cDNA
            clone pgm2n.pk008.c22 5' similar to ref|NP_060186.1 (NM_017716)
            membrane-spanning 4-domains, subfamily A, member 12 4-domains,
            subfamily A, member 7 [Homo sapiens] dbj|BAA91020.1| (AK000224)
            MS4A12 protein [Homo sapiens], mRNA sequence.
ACCESSION   BM488591
VERSION     BM488591.1  GI:18609522
KEYWORDS    EST.
SOURCE      chicken.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 687)
  AUTHORS   Cogburn,L.A. and Monsonego-Ornan,E.
  TITLE     ESTs from Normalized Chicken Breast Muscle, Leg Muscle, and
            Epiphyseal Growth Plate cDNA library, USDA/IFAFS Animal Genome
            Project
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Larry A. Cogburn
            University of Delaware
            Townsend Hall, Newark, DE 19717, USA
            Tel: 302-831-1335
            Fax: 302-831-2822
            Email: cogburn@udel.edu, www.chickest.udel.edu.
FEATURES             Location/Qualifiers
     source          1. .687
                     /organism="Gallus gallus"
                     /strain="Commercial broiler and  Ottawa Res. Centre
                     Strains 90 & 21"
                     /db_xref="taxon:9031"
                     /clone="pgm2n.pk008.c22"
                     /clone_lib="Normalized Chicken Breast Muscle, Leg Muscle,
                     and Epiphyseal Growth Plate cDNA library (pgm2n)"
                     /sex="Male and Female"
                     /tissue_type="Breast muscle, leg muscle and epiphyseal
                     growth plate"
                     /dev_stage="Breast,leg:Embryo(d19);post-hatch(1d,1,3,5,7,9
                     ,11 weeks);growth plate(1d,7d,14d post-hatch)"
                     /lab_host="E. coli EMDH10B"
                     /note="Vector: pCMVSPORT6; Library made from equivalent
                     pools of total RNA isolated from each tissue (embryonic
                     muscle 33.3%; juvenile muscle 33.3%; and epiphyseal growth
                     plate 33.3% of the final RNA pool). Single pass sequencing
                     from 5'-end"
BASE COUNT      129 a    223 c    188 g    143 t      4 others
ORIGIN

  Query Match           5.9%;  Score 77.6;  DB 13;  Length 687;
  Best Local Similarity   52.6%;  Pred. No. 4.1e-06;
  Matches  241;  Conservative    0;  Mismatches  184;  Indels   33;  Gaps    2;

Qy        492 TAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGGTTTTGGAATTGTTTTGTGTT 551
              | |||   ||||||||||||   || || ||||||||||| || || || ||  ||
Db        108 TGGGGACTATCCAGATCATGACCGGCTTCATGCACATTGGCTTCGGTATCGTCCTGACCA 167

Qy        552 TAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCTACTGCTGTTATTGGTGGAT 611
              | |  ||                      |  |   |||| |      | |  || |
Db        168 CACTCACC-----------------AATGTCTACTCCTCCATCTTCATCACCGGCGAGA 209

Qy        612 ACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCTCTCTCTGTGTCAGCATCCA 671
              ||| |||  ||| ||  | ||||| || ||||| |||  || ||  |    |||   |
Db        210 TCCCTTTCCTGGGCGGTGTGTCTTTCATCATCTCGGGCTGCCTGTCCATTGGGGCAGAGA 269

Qy        672 AGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATGAACATTGTTAGTTCTATCT 731
              ||   |   |  |||  |||||| ||||||| ||  | ||||||  | ||| | ||||
Db        270 AGAGTCCCACAGAGTGTGCGGTGAAGGGCAGCCAGGCCACGAACATCATCAGTGCCATCT 329

Qy        732 TGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGCATCAATGGGGTAGCTGG-- 789
              | ||  || | ||  |  |     |  |  | ||  |   | |||| ||| |     |
Db        330 TTGCGCTCCTGGGCATCGTCGCCTTCATCATTGACCTCAACTTCAACGGGCTGTACCGCT 389

Qy        790 ------------CCAAGACTACTGGGCCGTGCTTTCTGGAAAAGGCATTTCAGCCACGC 836
                         | |   ||||   | | ||||  | || || ||||| ||  |  ||
Db        390 CCAGCGATGACTACTACAGCTACCTTGTCCTGCTCGCAGGGAATGGCATCTCCATCGTGC 449

Qy        837 TGATGATCTTCTCCCTCTTGGAGTTCTTCGTAGCTTGTGCCACAGCCCATTTTGCCAACC 896
              || | |||||| || || |||||||||| | | |||   ||||| ||| | ||      ||
Db        450 TGCTCATCTTCACCATCCTGGAGTTCTGCATTGCTGTCGCCACTGCCAACTTCTGGTGCC 509

Qy        897 AAGCAAACACCACAACCAATATGTCTGTCCTGGTTATT 934
                |||| |  | | |  |  | | || | |||
Db        510 GGGCAACACGCCTCAGCCCCAACGAGGCCATGCTGATT 547


RESULT 10
BB629370
LOCUS       BB629370                673 bp    mRNA    linear   EST 31-AUG-2001
DEFINITION  BB629370 RIKEN full-length enriched, adult male bone Mus musculus
            cDNA clone 9830113C01 5', mRNA sequence.
ACCESSION   BB629370
VERSION     BB629370.1  GI:15399617
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 673)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Yamanaka,I., Kiyosawa,H., Kondo,S., Saito,T., Shinagawa,A., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K., Arakawa,T.,
            Ishii,Y. and Hayashizaki,Y.
             Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct.
            Func. Genomics 2 pre, L72-L86 (2001
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
```

```
          e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .673
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="9830113C01"
                     /clone_lib="RIKEN full-length enriched, adult male bone"
                     /sex="male"
                     /tissue_type="bone"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 185.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                     3']. cDNA was cleaved with XhoI and BamHI. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I."
BASE COUNT      146 a    158 c    167 g    202 t
ORIGIN

  Query Match             5.7%;  Score 75.6;  DB 10;  Length 673;
  Best Local Similarity   52.5%;  Pred. No. 1.1e-05;
  Matches  165;  Conservative    0;  Mismatches  149;  Indels    0;  Gaps    0;

Qy        486 AGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGGTTTTGGAATTGTTT 545
              | || || |||| |||||||||| ||  |||| |  ||    | | | | |||||| ||
Db        284 AAGCCCTCGGGGCCATCCAGATCCTGAATGGAATACTGATTCTAGCTCTCGGAATTTTTC 343

Qy        546 TGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCTACTGCTGTTATTG 605
              ||  ||   ||     |   |      | |      |||   ||     |  |
Db        344 TGGTTTGTTTACAACACGTGTCCCACCACTTCAGGCATTTCTTCTTCTTCACCTTCTACA 403

Qy        606 GTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCTCTCTCTGTGTCAG 665
                || |||||| | |||||||   | | |||||| | ||| || ||  |  |||| | |
Db        404 CAGGCTACCCACTGTGGGGTGCTGTGTTTTTTATCAGCTCAGGATCCTTGACTGTTGCCG 463

Qy        666 CATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATGAACATTGTTAGTT 725
              ||   |  | |   | ||    ||| || ||  |||  | || || |||||||  |||
Db        464 CAGGGAGAAACCCCACACGAATGCTGATGCAAAACAGTTTTGGGATAAACATTGCCAGTA 523

Qy        726 CTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGCATCAATGGGGTAG 785
              |||   | || ||  ||||     || | ||    ||| || ||    |||||      |
Db        524 CTACAATTGCATTTGTTGGGACTGTTTTCCTTTCTGTGCATTTGGCATTCAATACCCAGG 583

Qy        786 CTGGCCAAGACTAC 799
              ||  | | |  | |
Db        584 CTTTCAAGGGTTGC 597


RESULT 11
BG434337
LOCUS       BG434337                515 bp    mRNA    linear   EST 14-MAR-2001
DEFINITION  602506331F1 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4603590 5',
            mRNA sequence.
ACCESSION   BG434337
VERSION     BG434337.1  GI:13340843
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 515)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1337  row: n  column: 07
            High quality sequence stop: 515.
FEATURES             Location/Qualifiers
     source          1. .515
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4603590"
                     /clone_lib="NIH_MGC_79"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: placenta; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                     ); 5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).  Average
                     insert size 1.3 kb (range 0.5-4.0 kb). 15/15 colonies
                     contained inserts  by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
                     Library."
BASE COUNT      139 a    133 c     99 g    144 t
ORIGIN

  Query Match             5.7%;  Score 74.8;  DB 12;  Length 515;
  Best Local Similarity   55.3%;  Pred. No. 1.6e-05;
  Matches  167;  Conservative    0;  Mismatches  132;  Indels    3;  Gaps    1;

Qy        474 AAGAAGAAGCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGGTT 533
              | | ||||   || |   | |||   |||||||||| ||  |||  |||||    |  | |
Db        147 ACGCAGAAATCAAAGTTATTGGGACTATCCAGATCTTGTGTGGCATGATGGTATTGAGCT 206

Qy        534 TTGGAATTGTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCTA 593
              | || ||  |||||   |     |||||||||     | |  |        |  | ||||
Db        207 TGGGGATCATTTTGGCATCTGCTTCCTTCTCTC---CAAATTTTACCCAAGTGACTTCTA 263

Qy        594 CTGCTGTTATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCTC 653
              |     | |    ||  |||||||||   ||   | ||| ||||||  ||||||||||||
Db        264 CACTGTTGAACTCTGCTTACCCATTCATAGGACCCTTTTTTTTTATCATCTCTGGCTCTC 323

Qy        654 TCTCTGTGTCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATGA 713
              | ||  |  |  ||   ||   | |  ||     | |||||  |  ||||||||     |
Db        324 TATCAATCGCCACAGAGAAAAGGTTAACCAAGCTTTTGGTGCATAGCAGCCTGGTTGGAA 383

Qy        714 ACATTGTTAGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGCA 773
              |||| | ||| || | |  ||| |  | ||  | ||| | |||   ||  |   |  ||
Db        384 GCATTCTGAGTGCTCTGTCTGCCCTGGTGGGTTTCATTATCCTGTCTGTCAAACAGGCCA 443

Qy        774 TC 775
              |
Db        444 CC 445


RESULT 12
AV717594
LOCUS       AV717594                594 bp    mRNA    linear   EST 16-OCT-2000
DEFINITION  AV717594 DCB Homo sapiens cDNA clone DCBCFE05 5', mRNA sequence.
ACCESSION   AV717594
VERSION     AV717594.1  GI:10814746
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 594)
  AUTHORS   Xu,X., Gu,J., Liu,F., Qu,J., Zhao,M., Li,Y., Huang,Q., Zhou,J.,
            Song,H., Gu,Y., Yang,Y., Gao,G., Xiao,H., Li,N., Qian,B., Gao,X.,
            Cheng,Z., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu
            ,G., Cheng,Z. and Han,Z.
  TITLE     Homo sapiens cDNA DCB clones
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1..594
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DCBCFE05"
                     /clone_lib="DCB"
                     /cell_type="dendritic cells"
                     /dev_stage="mature"
                     /lab_host="BM25.8"
                     /note="Vector: pTriplEx2; Site_1: sfiIA; Site_2: sfiIB"
BASE COUNT      162 a    150 c    124 g    158 t
ORIGIN

  Query Match             5.7%;  Score 74.8;  DB 10;  Length 594;
  Best Local Similarity   55.3%;  Pred. No. 1.6e-05;
  Matches  167;  Conservative    0;  Mismatches  132;  Indels    3;  Gaps    1;

Qy        474 AAGAAGAAGCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGGTT 533
              | | ||||   || |   | |||   ||||||||| ||  |||  |||||    |  | |
Db        244 ACGCAGAAATCAAAGTTATTGGGACTATCCAGATCTTGTGTGGCATGATGGTATTGAGCT 303

Qy        534 TTGGAATTGTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCTA 593
              | || ||  |||||   |     |||||||||     | |  |        |  | ||||
Db        304 TGGGGATCATTTTGGCATCTGCTTCCTTCTCTC---CAAATTTTACCCAAGTGACTTCTA 360

Qy        594 CTGCTGTTATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCTC 653
              |     | |    ||  |||||||||    ||   | ||| ||||||  ||||||||||||
Db        361 CACTGTTGAACTCTGCTTACCCATTCATAGGACCCTTTTTTTTTTATCATCTCTGGCTCTC 420

Qy        654 TCTCTGTGTCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATGA 713
              | ||  |  |  ||   ||   | |  ||     | ||||| |  |||||||||     |
Db        421 TATCAATCGCCACAGAGAAAAGGTTAACCAAGCTTTTGGTGCATAGCAGCCTGGTTGGAA 480

Qy        714 ACATTGTTAGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGCA 773
               |||| | ||| || | |  ||| |  | ||  | ||| | |||   ||  |   |  ||
Db        481 GCATTCTGAGTGCTCTGTCTGCCCTGGTGGGTTTCATTATCCTGTCTGTCAAACAGGCCA 540

Qy        774 TC 775
              |
Db        541 CC 542


RESULT 13
AV661743
LOCUS       AV661743                619 bp    mRNA    linear   EST 16-JAN-2002
DEFINITION  AV661743 GLC Homo sapiens cDNA clone GLCGXB07 3', mRNA sequence.
ACCESSION   AV661743
VERSION     AV661743.1  GI:9882757
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 619)
  AUTHORS   Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
            Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
            Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
            Hu,G., Gu,J., Chen,Z. and Han,Z.
  TITLE     Insight into hepatocellular carcinogenesis at transcriptome level
            by comparing gene expression profiles of hepatocellular carcinoma
            with those of corresponding noncancerous liver
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
  MEDLINE   21625106
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1..619
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="GLCGXB07"
                     /clone_lib="GLC"
                     /tissue_type="corresponding non cancerous liver tissue"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      153 a    162 c    133 g    171 t
ORIGIN

  Query Match             5.7%;  Score 74.8;  DB 10;  Length 619;
  Best Local Similarity   55.3%;  Pred. No. 1.6e-05;
  Matches  167;  Conservative    0;  Mismatches  132;  Indels    3;  Gaps    1;

Qy        474 AAGAAGAAGCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGGTT 533
              | | ||||   || |   | |||   ||||||||| ||  |||  |||||    |  | |
Db        177 ACGCAGAAATCAAAGTTATTGGGACTATCCAGATCTTGTGTGGCATGATGGTATTGAGCT 236

Qy        534 TTGGAATTGTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCTA 593
              | || ||  |||||   |     |||||||||   |          ||   |  | ||||
Db        237 TGGGGATCATTTTGGCATCTGCTTCCTTCTCTCCAAATTTTACCCAAG---TGACTTCTA 293

Qy        594 CTGCTGTTATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCTC 653
              |     | |    ||  |||||||||    ||   | ||| ||||||  ||||||||||||
Db        294 CACTGTTGAACTCTGCTTACCCATTCATAGGACCCTTTTTTTTTATCATCTCTGGCTCTC 353

Qy        654 TCTCTGTGTCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATGA 713
              | ||  |  |  ||   ||   | |  ||     | ||||| |  |||||||||     |
Db        354 TATCAATCGCCACAGAGAAAAGGTTAACCAAGCTTTTGGTGCATAGCAGCCTGGTTGGAA 413

Qy        714 ACATTGTTAGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGCA 773
               |||| | ||| || | |  ||| |  | ||  | ||| | |||   ||  |   |  ||
Db        414 GCATTCTGAGTGCTCTGTCTGCCCTGGTGGGTTTCATTATCCTGTCTGTCAAACAGGCCA 473

Qy        774 TC 775
              |
Db        474 CC 475


RESULT 14
BG546745
LOCUS       BG546745                680 bp    mRNA    linear   EST 04-APR-2001
DEFINITION  602574115F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4702274 5',
            mRNA sequence.
ACCESSION   BG546745
VERSION     BG546745.1  GI:13545410
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 680)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
```

```
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1538  row: n  column: 03
            High quality sequence stop: 677.
FEATURES             Location/Qualifiers
     source          1. .680
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4702274"
                     /clone_lib="NIH_MGC_77"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: lung; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.9
                     kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
BASE COUNT      191 a    167 c    136 g    186 t
ORIGIN

  Query Match             5.7%;  Score 74.8;  DB 12;  Length 680;
  Best Local Similarity   55.3%;  Pred. No. 1.6e-05;
  Matches  167;  Conservative    0;  Mismatches  132;  Indels    3;  Gaps    1;

Qy        474 AAGAAGAAGCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGGTT 533
              | | ||||   || |   | |||  ||||||||| ||  |||  |||||    |  | |
Db        246 ACGCAGAAATCAAAGTTATTGGGACTATCCAGATCTTGTGTGGCATGATGGTATTGAGCT 305

Qy        534 TTGGAATTGTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCTA 593
              | || ||  |||||   |     |||||||||   |          ||   |  | ||||
Db        306 TGGGGATCATTTTGGCATCTGCTTCCTTCTCTCCAAATTTTACCCAAG---TGACTTCTA 362

Qy        594 CTGCTGTTATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCTC 653
              |     | |    ||  |||||||||   ||   | ||| |||||| ||||||||||||
Db        363 CACTGTTGAACTCTGCTTACCCATTCATAGGACCCTTTTTTTTTTATCATCTCTGGCTCTC 422

Qy        654 TCTCTGTGTCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATGA 713
              | ||  |  |  ||   ||   | |  ||     | |||||  |  |||||||||     |
Db        423 TATCAATCGCCACAGAGAAAAGGTTAACCAAGCTTTTGGTGCATAGCAGCCTGGTTGGAA 482

Qy        714 ACATTGTTAGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGCA 773
              |||| | ||| || | |  ||| | | ||  | ||| | |||   ||  |   |  ||
Db        483 GCATTCTGAGTGCTCTGTCTGCCCTGGTGGGTTTCATTATCCTGTCTGTCAAACAGGCCA 542

Qy        774 TC 775
              |
Db        543 CC 544


RESULT 15
AV715678
LOCUS       AV715678                735 bp    mRNA    linear   EST 11-OCT-2000
DEFINITION  AV715678 DCB Homo sapiens cDNA clone DCBAXG06 5', mRNA sequence.
ACCESSION   AV715678
VERSION     AV715678.1  GI:10797195
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 735)
  AUTHORS   Xu,X., Gu,J., Liu,F., Qu,J., Zhao,M., Li,Y., Huang,Q., Zhou,J.,
            Song,H., Gu,Y., Yang,Y., Gao,G., Xiao,H., Li,N., Qian,B., Gao,X.,
            Cheng,Z., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu
            ,G., Cheng,Z. and Han,Z.
  TITLE     Homo sapiens cDNA DCB clones
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1. .735
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DCBAXG06"
                     /clone_lib="DCB"
                     /cell_type="dendritic cells"
                     /dev_stage="mature"
                     /lab_host="BM25.8"
                     /note="Vector: pTriplEx2; Site_1: sfiIA; Site_2: sfiIB"
BASE COUNT      193 a    183 c    154 g    204 t      1 others
ORIGIN

  Query Match             5.7%;  Score 74.8;  DB 10;  Length 735;
  Best Local Similarity   55.3%;  Pred. No. 1.6e-05;
  Matches  167;  Conservative    0;  Mismatches  132;  Indels    3;  Gaps    1;

Qy        474 AAGAAGAAGCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGGTT 533
              | | ||||   || |   | |||  ||||||||| ||  |||  |||||    |  | |
Db        236 ACGCAGAAATCAAAGTTATTGGGACTATCCAGATCTTGTGTGGCATGATGGTATTGAGCT 295

Qy        534 TTGGAATTGTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCTA 593
              | || ||  |||||   |     |||||||||   |          ||   |  | ||||
Db        296 TGGGGATCATTTTGGCATCTGCTTCCTTCTCTCCAAATTTTACCCAAG---TGACTTCTA 352

Qy        594 CTGCTGTTATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCTC 653
              |     | |    ||  |||||||||   ||   | ||| |||||| ||||||||||||
Db        353 CACTGTTGAACTCTGCTTACCCATTCATAGGACCCTTTTTTTTTTATCATCTCTGGCTCTC 412

Qy        654 TCTCTGTGTCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATGA 713
              | ||  |  |  ||   ||   | |  ||     | |||||  |  |||||||||     |
Db        413 TATCAATCGCCACAGAGAAAAGGTTAACCAAGCTTTTGGTGCATAGCAGCCTGGTTGGAA 472

Qy        714 ACATTGTTAGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGCA 773
              |||| | ||| || | |  ||| | | ||  | ||| | |||   ||  |   |  ||
Db        473 GCATTCTGAGTGCTCTGTCTGCCCTGGTGGGTTTCATTATCCTGTCTGTCAAACAGGCCA 532

Qy        774 TC 775
              |
Db        533 CC 534


Search completed: July 16, 2003, 14:55:39
Job time : 1441.76 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 16, 2003, 10:35:27 ; Search time 51.3795 Seconds
(without alignments)
7860.974 Million cell updates/sec

Title: US-09-981-353-81
Perfect score: 1317
Sequence: 1 aaggaacaaagtaagtccat.....tttttccttaataaaatgtc 1317

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	74.8	5.7	945	4	US-09-149-476-168	Sequence 168, App
	2	71.2	5.4	2545	1	US-07-869-933-22	Sequence 22, Appl
	3	71.2	5.4	2545	4	US-09-103-663-22	Sequence 22, Appl
	4	53	4.0	1669	2	US-08-916-902A-2	Sequence 2, Appli
	5	53	4.0	1669	2	US-09-213-389-2	Sequence 2, Appli
	6	51.6	3.9	1661	1	US-08-318-492-3	Sequence 3, Appli
	7	51.6	3.9	1661	1	US-08-707-340-3	Sequence 3, Appli
	8	51.6	3.9	1661	2	US-08-994-578-3	Sequence 3, Appli
c	9	51.6	3.9	7218	1	US-08-232-463-14	Sequence 14, Appl
	10	49.4	3.8	1308	4	US-09-724-864-15	Sequence 15, Appl
	11	44.4	3.4	7218	1	US-08-232-463-14	Sequence 14, Appl
	12	40.8	3.1	2621	2	US-08-553-619B-8	Sequence 8, Appli
	13	40.2	3.1	2065	4	US-09-370-473-5	Sequence 5, Appli
	14	38.6	2.9	1798	4	US-09-797-906-1	Sequence 1, Appli
	15	38.4	2.9	2207	6	5221620-3	Patent No. 5221620
	16	38.4	2.9	2569	6	5221620-1	Patent No. 5221620
	17	38.2	2.9	291	1	US-07-922-723A-7	Sequence 7, Appli
	18	38.2	2.9	291	1	US-07-799-828C-7	Sequence 7, Appli
	19	38.2	2.9	291	1	US-08-074-275-7	Sequence 7, Appli
	20	38.2	2.9	291	1	US-08-480-366-7	Sequence 7, Appli
	21	38.2	2.9	291	2	US-07-952-277A-7	Sequence 7, Appli
	22	38.2	2.9	152331	3	US-09-128-155-16	Sequence 16, Appl
	23	37.6	2.9	454	2	US-08-623-906A-6	Sequence 6, Appli
c	24	37.2	2.8	227	2	US-08-520-678A-28	Sequence 28, Appl
c	25	37.2	2.8	227	4	US-08-897-126-28	Sequence 28, Appl
	26	37.2	2.8	2674	4	US-09-817-180-1	Sequence 1, Appli
	27	36.8	2.8	2233	1	US-08-496-631-1	Sequence 1, Appli
	28	36.8	2.8	6638	2	US-08-070-301-2	Sequence 2, Appli
c	29	36.6	2.8	6755	3	US-08-931-999-4	Sequence 4, Appli
c	30	36.6	2.8	176373	3	US-09-128-155-17	Sequence 17, Appl
	31	36	2.7	1066	1	US-08-157-101A-4	Sequence 4, Appli
	32	36	2.7	1069	4	US-09-372-422A-7	Sequence 7, Appli
	33	36	2.7	3275	4	US-09-370-838-151	Sequence 151, App
	34	35.8	2.7	2447	2	US-09-014-969-14	Sequence 14, Appl
c	35	35.6	2.7	356	2	US-08-520-678A-22	Sequence 22, Appl
c	36	35.6	2.7	356	4	US-08-897-126-22	Sequence 22, Appl
	37	35.6	2.7	376	2	US-08-623-906A-18	Sequence 18, Appl
c	38	35.6	2.7	9646	3	US-08-811-566-1	Sequence 1, Appli
c	39	35.6	2.7	9646	4	US-09-034-756-1	Sequence 1, Appli
	40	35.4	2.7	1582	3	US-08-545-196B-10	Sequence 10, Appl
	41	35.4	2.7	1582	3	US-08-545-196B-12	Sequence 12, Appl
	42	35.4	2.7	2873	4	US-08-630-915A-193	Sequence 193, App
c	43	35.4	2.7	12980	3	US-08-811-566-5	Sequence 5, Appli
c	44	35.4	2.7	12980	4	US-09-034-756-5	Sequence 5, Appli
	45	35.2	2.7	1117	4	US-09-247-373B-33	Sequence 33, Appl

ALIGNMENTS

RESULT 1
US-09-149-476-168
; Sequence 168, Application US/09149476
; Patent No. 6420526
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: PZ002P1
; CURRENT APPLICATION NUMBER: US/09/149,476
; CURRENT FILING DATE: 1998-09-08
; EARLIER APPLICATION NUMBER: PCT/US98/04493
; EARLIER FILING DATE: 1998-03-06
; EARLIER APPLICATION NUMBER: 60/040,162
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,333
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/038,621
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,626
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,334
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,336
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,163
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/047,600
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,615
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,597
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,502
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,633
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,583
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,617
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,618
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,503
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,592
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,581
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,584
; EARLIER FILING DATE: 1997-05-23

; EARLIER APPLICATION NUMBER: 60/047,500
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,587
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,492
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,598
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,613
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,582
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,596
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,612
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,632
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,601
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,580
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,568
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,314
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,569
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,311
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,671
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,674
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,669
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,312
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,313
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,672
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,315
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/048,974
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/056,886
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,877
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,889
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,893
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,630
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,878
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,662
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,872
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,882
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,637
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,903
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,888
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,879
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,880
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,894
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,911
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,636
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,874
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,910
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,864
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,631
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,845
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,892
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/057,761
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/047,595
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,599
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,588
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,585
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,586
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,590
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,594
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,589
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,593
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,614
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,578
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,576
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/047,501
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,670
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/056,632
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,664
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,876
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,881
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,909
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,875
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,862
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,887
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,908
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/048,964
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/057,650
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: 60/056,884
; EARLIER FILING DATE: 1997-08-22

```
;  EARLIER APPLICATION NUMBER: 60/057,669
;  EARLIER FILING DATE: 1997-09-05
;  EARLIER APPLICATION NUMBER: 60/049,610
;  EARLIER FILING DATE: 1997-06-13
;  EARLIER APPLICATION NUMBER: 60/061,060
;  EARLIER FILING DATE: 1997-10-02

  Query Match             5.7%;  Score 74.8;  DB 4;  Length 945;
  Best Local Similarity   55.3%;  Pred. No. 7.8e-12;
  Matches  167;  Conservative    0;  Mismatches  132;  Indels    3;  Gaps    1;

Qy        474 AAGAAGAAGCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGGTT 533
              | | ||||   || |   | |||   ||||||||| ||  |||  |||||    |  | |
Db        164 ACGCAGAAATCAAAGTTATTGGGACTATCCAGATCTTGTGTGGCATGATGGTATTGAGCT 223

Qy        534 TTGGAATTGTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCTA 593
              | || ||  |||||   |     |||||||||   |          ||   |  | ||||
Db        224 TGGGGATCATTTTGGCATCTGCTTCCTTCTCTCCAAATTTTACCCAAG---TGACTTCTA 280

Qy        594 CTGCTGTTATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCTC 653
              |     | |    ||  |||||||||   ||   | ||| |||||| ||||||||||||
Db        281 CACTGTTGAACTCTGCTTACCCATTCATAGGACCCTTTTTTTTTATCATCTCTGGCTCTC 340

Qy        654 TCTCTGTGTCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATGA 713
              | ||  |  |  ||   ||  | |  ||     | ||||| |  |||||||||     |
Db        341 TATCAATCGCCACAGAGAAAAGGTTRACCAAGCTTTTGGTGCATAGCAGCCTGGTTGGAA 400

Qy        714 ACATTGTTAGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGCA 773
              |||| | ||| || | |  ||| |  | ||  | ||| | |||   ||  |   |  ||
Db        401 GCATTCTGAGTGCTCTGTCTGCCCTGGTGGGTTTCATTATCCTGTCTGTCAAACAGGCCA 460

Qy        774 TC 775
              |
Db        461 CC 462


RESULT 2
US-07-869-933-22
; Sequence 22, Application US/07869933
; Patent No. 5770396
;  GENERAL INFORMATION:
;    APPLICANT:  KINET, Jean-Pierre
;    TITLE OF INVENTION:  ISOLATION, CHARACTERIZATION, AND USE OF
;    TITLE OF INVENTION:  THE HUMAN B SUBUNIT OF THE HIGH AFFINITY RECEPTOR FOR
;    TITLE OF INVENTION:  IMMUNOGLOBULIN
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/869,933
;      FILING DATE:  19920416
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  40399/154 NIHD
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  22:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2545 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  46..786
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  46..54
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  55..786
US-07-869-933-22

  Query Match             5.4%;  Score 71.2;  DB 1;  Length 2545;
  Best Local Similarity   54.2%;  Pred. No. 1.4e-10;
  Matches  167;  Conservative    0;  Mismatches  138;  Indels    3;  Gaps    1;

Qy        470 TTTAAAGAAGAAGCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATT 529
              || ||  ||||      ||   || || || | |||  |  ||||||| |||||    | ||
Db        214 TTGAAGAAAGAGTTGGAGTTCCTGGGCGTAACCCAAGTTCTGGTTGGTTTGATATGCCTT 273

Qy        530 GGTTTTGGAATTGTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCC 589
               |||||||||  ||| | || |  | |  |    |||    |||  |     |      |
Db        274 TGTTTTGGAACAGTTGTCTGCTCCACACTCCAGACTTC---AGACTTTGACGACGAAGTG 330

Qy        590 TCTACTGCTGTTATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGC 649
                |         ||  |  || ||||||||||||||||    |    ||| || | |||||
Db        331 CTTTTATTATATAGAGCAGGCTACCCATTCTGGGGTGCAGTGCTGTTTGTTTTGTCTGGA 390

Qy        650 TCTCTCTCTGTGTCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGA 709
              | | | ||  |      |    | | |     | |  | |||||||| ||||||||||||
Db        391 TTTTTGTCAATTATGTCCGAAAGGAAAAACACACTGTATCTGGTGAGAGGCAGCCTGGGA 450

Qy        710 ATGAACATTGTTAGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATG 769
                |||||||| ||    |||   ||   | | ||  |      |  || |  |  || ||
Db        451 GCAAACATTGTCAGCAGCATCGCTGCAGGCTTGGGGATCGCCATATTGATTCTCAATCTG 510

Qy        770 TGCATCAA 777
                ||| |||
Db        511 AGCAACAA 518


RESULT 3
US-09-103-663-22
; Sequence 22, Application US/09103663D
; Patent No. 6171803
; GENERAL INFORMATION:
;  APPLICANT: Kinet et al.
;  TITLE OF INVENTION: Isolation, characterization, and use of the human beta
;  TITLE OF INVENTION: subunit of the high affinity receptor for
;  TITLE OF INVENTION: immunoglobulin E.
;  FILE REFERENCE: 50490
;  CURRENT APPLICATION NUMBER: US/09/103,663D
;  CURRENT FILING DATE: 1998-06-23
;  EARLIER APPLICATION NUMBER: 07/869,933
;  EARLIER FILING DATE: 1992-04-16
;  NUMBER OF SEQ ID NOS: 35
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
;   LENGTH: 2545
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: sig_peptide
;   LOCATION: (46)..(54)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (46)..(786)
US-09-103-663-22

  Query Match             5.4%;  Score 71.2;  DB 4;  Length 2545;
```

```
  Best Local Similarity   54.2%;  Pred. No. 1.4e-10;
  Matches  167;  Conservative    0;  Mismatches  138;  Indels    3;  Gaps    1;

Qy        470 TTTAAAGAAGAAGCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATT 529
              || ||  ||||     ||  || || || | |||  |  ||||||| ||||  | ||
Db        214 TTGAAGAAAGAGTTGGAGTTCCTGGGCGTAACCCAAGTTCTGGTTGGTTTGATATGCCTT 273

Qy        530 GGTTTTGGAATTGTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCC 589
              |||||||||  ||| | || |  | |  |    |||    |||  |    |      |
Db        274 TGTTTTGGAACAGTTGTCTGCTCCACACTCCAGACTTC---AGACTTTGACGACGAAGTG 330

Qy        590 TCTACTGCTGTTATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGC 649
              |       ||  |  || ||||||||||||||||   |    ||| || | |||||
Db        331 CTTTTATTATATAGAGCAGGCTACCCATTCTGGGGTGCAGTGCTGTTTGTTTTGTCTGGA 390

Qy        650 TCTCTCTCTGTGTCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGA 709
              | | | ||  |     |    | | |     | |  | ||||||| ||||||||||||
Db        391 TTTTTGTCAATTATGTCCGAAAGGAAAAACACACTGTATCTGGTGAGAGGCAGCCTGGGA 450

Qy        710 ATGAACATTGTTAGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATG 769
              ||||||| ||    |||   ||   | | ||  |     |  || |  |  || ||
Db        451 GCAAACATTGTCAGCAGCATCGCTGCAGGCTTGGGGATCGCCATATTGATTCTCAATCTG 510

Qy        770 TGCATCAA 777
              ||| |||
Db        511 AGCAACAA 518


RESULT 4
US-08-916-902A-2
; Sequence 2, Application US/08916902A
; Patent No. 5871930
;  GENERAL INFORMATION:
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Corley, Neil C.
;    TITLE OF INVENTION:  HIGH AFFINITY IMMUNOGLOBULIN E
;    TITLE OF INVENTION:  RECEPTOR-LIKE PROTEIN
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/916,902A
;      FILING DATE:  Herewith
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0371 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1669 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  BRAINOT04
;      CLONE:  927955
US-08-916-902A-2

  Query Match             4.0%;  Score 53;  DB 2;  Length 1669;
  Best Local Similarity   51.5%;  Pred. No. 1.9e-05;
  Matches  153;  Conservative    0;  Mismatches  135;  Indels    9;  Gaps    1;

Qy        439 AAATCCAAGTGTGGGAACAGCAGTAATGAACTTTAAAGAAGAAGCAAAGGCACTAGGGGT 498
              | |||   |    |||     ||  | |  ||| || | |||| | || |  || |||||
Db        302 ACATCTGTGGAAAGGATTGCAAGAGAAGTTCTTGAAGGGAGAACCCAAAGTCCTTGGGGT 361

Qy        499 GATCCAGATCATGGTTGGATTGATGCACATTGGTTTTGGAATTGTTTTGTGTTTAATATC 558
               | |||||  ||  ||   |||||  | || |  | |||||     ||   |  ||  |
Db        362 TGTGCAGATTCTGACTGCCCTGATGAGCCTTAGCATGGGAATAACAATGATGTGTATGGC 421

Qy        559 CTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCTACTGCTGTTATTGGTGGATACCCATT 618
              |    | |          |  ||   |  | ||| | | ||   |   || ||| || |
Db        422 ATCTAATACT----------TATGGAAGTAACCCTATTTCCGTGTATATCGGGTACACAAT 472

Qy        619 CTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCTCTCTCTGTGTCAGCATCCAAGGAGCT 678
              |||||      |    |||||||| || || ||  | ||  |  |||||    |
Db        473 TTGGGGGTCAGTAATGTTTATTATTTCAGGATCCTTGTCAATTGCAGCAGGAATTAGAAC 532

Qy        679 TTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATGAACATTGTTAGTTCTATCTTGGC 735
              | |      | |||||   ||| || || |||||||| ||    || ||| |  ||||
Db        533 TACAAAAGGCCTGGTCCGAGGTAGTCTAGGAATGAATATCACCAGCTCTGTACTGGC 589


RESULT 5
US-09-213-389-2
; Sequence 2, Application US/09213389
; Patent No. 5977072
;  GENERAL INFORMATION:
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Corley, Neil C.
;    TITLE OF INVENTION:  HIGH AFFINITY IMMUNOGLOBULIN E
;    TITLE OF INVENTION:  RECEPTOR-LIKE PROTEIN
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/213,389
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/916,902
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0371 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1669 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
```

```
;    IMMEDIATE SOURCE:
;      LIBRARY:  BRAINOT04
;      CLONE:  927955
US-09-213-389-2

  Query Match             4.0%;  Score 53;  DB 2;  Length 1669;
  Best Local Similarity   51.5%;  Pred. No. 1.9e-05;
  Matches  153;  Conservative    0;  Mismatches  135;  Indels    9;  Gaps    1;

Qy        439 AAATCCAAGTGTGGGAACAGCAGTAATGAACTTTAAAGAAGAAGCAAAGGCACTAGGGGT 498
              | |||   |   |||    ||  | |  ||| || | |||| | || |  || |||||
Db        302 ACATCTGTGGAAAGGATTGCAAGAGAAGTTCTTGAAGGGAGAACCCAAAGTCCTTGGGGT 361

Qy        499 GATCCAGATCATGGTTGGATTGATGCACATTGGTTTTGGAATTGTTTTGTGTTTAATATC 558
              | |||||  ||  ||   |||||  | || |  | |||||     ||   |  ||  |
Db        362 TGTGCAGATTCTGACTGCCCTGATGAGCCTTAGCATGGGAATAACAATGATGTGTATGGC 421

Qy        559 CTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCTACTGCTGTTATTGGTGGATACCCATT 618
              |    | |         |  ||   | | ||| | | ||   |   || ||| || |
Db        422 ATCTAATACT---------TATGGAAGTAACCCTATTTCCGTGTATATCGGGTACACAAT 472

Qy        619 CTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCTCTCTCTGTGTCAGCATCCAAGGAGCT 678
              |||||     |   |||||||| || || ||  | ||  |  |||||   |
Db        473 TTGGGGGTCAGTAATGTTTATTATTTCAGGATCCTTGTCAATTGCAGCAGGAATTAGAAC 532

Qy        679 TTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATGAACATTGTTAGTTCTATCTTGGC 735
              | |     | |||||   ||| || || |||||||| ||     || ||| |  ||||
Db        533 TACAAAAGGCCTGGTCCGAGGTAGTCTAGGAATGAATATCACCAGCTCTGTACTGGC 589


RESULT 6
US-08-318-492-3
; Sequence 3, Application US/08318492
; Patent No. 5552312
;  GENERAL INFORMATION:
;    APPLICANT:  Lim, Bing
;    APPLICANT:  Adra, Chaker N.
;    APPLICANT:  Lelias, Jean-Michel
;    TITLE OF INVENTION:  RECOMBINANT HTM4 GENE, PROTEIN AND
;    TITLE OF INVENTION:  ASSAYS
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/318,492
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brook, David E.
;      REGISTRATION NUMBER:  22,592
;      REFERENCE/DOCKET NUMBER:  BIH94-03
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 861-6240
;      TELEFAX:  (617) 861-9540
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1661 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  97..741
US-08-318-492-3

  Query Match             3.9%;  Score 51.6;  DB 1;  Length 1661;
  Best Local Similarity   48.0%;  Pred. No. 4.9e-05;
  Matches  147;  Conservative    0;  Mismatches  159;  Indels    0;  Gaps    0;

Qy        473 AAAGAAGAAGCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGGT 532
              |||| | ||  | | |  || ||||  |||||||| ||  || |   |||    | | |
Db        229 AAAGCAAAATTACAAGTTCTTGGGGCCATCCAGATCCTGAATGCAGCAATGATTCTGGCT 288

Qy        533 TTTGGAATTGTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCT 592
              || ||  |  || || |||   |    | | | |           |      ||    |
Db        289 TTGGGTGTCTTTCTGGGTTCCTTGCAATACCCATACCACTTCCAAAAGCACTTCTTTTTC 348

Qy        593 ACTGCTGTTATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCT 652
                |  |        || |||||  | |||||||   | | |||   ||  || ||  |
Db        349 TTCACCTTCTACACAGGCTACCCGATTTGGGGTGCTGTGTTTTTCTGTAGTTCAGGAACC 408

Qy        653 CTCTCTGTGTCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATG 712
              | |||||   ||||   |   | |   |  |      | |  |   |||  | ||||||
Db        409 TTGTCTGTTGTAGCAGGGATAAAACCCACAAGAACATGGATACAGAACAGTTTTGGAATG 468

Qy        713 AACATTGTTAGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGC 772
              |||||||  ||| |||   | ||  |  | ||      | | ||     | ||||
Db        469 AACATTGCCAGTGCTACAATTGCACTAGTGGGGACTGCTTTTCTCTCACTAAATATAGCA 528

Qy        773 ATCAAT 778
                | |||
Db        529 GTTAAT 534


RESULT 7
US-08-707-340-3
; Sequence 3, Application US/08707340
; Patent No. 5705615
;  GENERAL INFORMATION:
;    APPLICANT:  Lim, Bing
;    APPLICANT:  Adra, Chaker N.
;    APPLICANT:  Lelias, Jean-Michel
;    TITLE OF INVENTION:  RECOMBINANT HTM4 GENE, PROTEIN AND
;    TITLE OF INVENTION:  ASSAYS
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/707,340
;      FILING DATE:  03-SEP-1996
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/318,492
;      FILING DATE:  06-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/675,648
;      FILING DATE:  03-JUL-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brook, David E.
;      REGISTRATION NUMBER:  22,592
;      REFERENCE/DOCKET NUMBER:  BIH94-03A2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 861-6240
;      TELEFAX:  (617) 861-9540
;  INFORMATION FOR SEQ ID NO:  3:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1661 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  97..738
US-08-707-340-3

  Query Match             3.9%;  Score 51.6;  DB 1;  Length 1661;
  Best Local Similarity   48.0%;  Pred. No. 4.9e-05;
  Matches  147;  Conservative    0;  Mismatches  159;  Indels    0;  Gaps    0;

Qy        473 AAAGAAGAAGCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGGT 532
              |||| | ||  | | |  || ||||  |||||||||| ||  || |   |||    | | |
Db        229 AAAGCAAAATTACAAGTTCTTGGGGCCATCCAGATCCTGAATGCAGCAATGATTCTGGCT 288

Qy        533 TTTGGAATTGTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCT 592
              || ||  |  || || |||   |    | | | |          |       ||    |
Db        289 TTGGGTGTCTTTCTGGGTTCCTTGCAATACCCATACCACTTCCAAAAGCACTTCTTTTTC 348

Qy        593 ACTGCTGTTATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCT 652
                |  |       || |||||  | |||||||   | | |||   ||  || ||  |
Db        349 TTCACCTTCTACACAGGCTACCCGATTTGGGGTGCTGTGTTTTTCTGTAGTTCAGGAACC 408

Qy        653 CTCTCTGTGTCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATG 712
               | |||||   ||||   |  | |   |  |      | |  |   |||  | ||||||
Db        409 TTGTCTGTTGTAGCAGGGATAAAACCCACAAGAACATGGATACAGAACAGTTTTGGAATG 468

Qy        713 AACATTGTTAGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGC 772
              |||||||  ||| |||   | ||  |  | ||      | | ||     |  ||||
Db        469 AACATTGCCAGTGCTACAATTGCACTAGTGGGGACTGCTTTTCTCTCACTAAATATAGCA 528

Qy        773 ATCAAT 778
                | |||
Db        529 GTTAAT 534


RESULT 8
US-08-994-578-3
; Sequence 3, Application US/08994578
; Patent No. 5972688
;  GENERAL INFORMATION:
;    APPLICANT:  Lim, Bing
;    APPLICANT:  Adra, Chaker N.
;    APPLICANT:  Lelias, Jean-Michel
;    TITLE OF INVENTION:  HTm4 METHODS OF TREATMENT AND ASSAYS,
;    TITLE OF INVENTION:  AGONISTS AND ANTAGONISTS
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/994,578
;      FILING DATE:  December 19, 1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/707,340
;      FILING DATE:  03-SEP-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/675,648
;      FILING DATE:  03-JUL-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/318,492
;      FILING DATE:  06-OCT-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brook, David E.
;      REGISTRATION NUMBER:  22,592
;      REFERENCE/DOCKET NUMBER:  BIH94-03A2Z
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (781) 861-6240
;      TELEFAX:  (781) 861-9540
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1661 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  97..738
US-08-994-578-3

  Query Match             3.9%;  Score 51.6;  DB 2;  Length 1661;
  Best Local Similarity   48.0%;  Pred. No. 4.9e-05;
  Matches  147;  Conservative    0;  Mismatches  159;  Indels    0;  Gaps    0;

Qy        473 AAAGAAGAAGCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGGT 532
              |||| | ||  | | |  || ||||  |||||||||| ||  || |   |||    | | |
Db        229 AAAGCAAAATTACAAGTTCTTGGGGCCATCCAGATCCTGAATGCAGCAATGATTCTGGCT 288

Qy        533 TTTGGAATTGTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCT 592
              || ||  |  || || |||   |    | | | |          |       ||    |
Db        289 TTGGGTGTCTTTCTGGGTTCCTTGCAATACCCATACCACTTCCAAAAGCACTTCTTTTTC 348

Qy        593 ACTGCTGTTATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCT 652
                |  |       || |||||  | |||||||   | | |||   ||  || ||  |
Db        349 TTCACCTTCTACACAGGCTACCCGATTTGGGGTGCTGTGTTTTTCTGTAGTTCAGGAACC 408

Qy        653 CTCTCTGTGTCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATG 712
               | |||||   ||||   |  | |   |  |      | |  |   |||  | ||||||
Db        409 TTGTCTGTTGTAGCAGGGATAAAACCCACAAGAACATGGATACAGAACAGTTTTGGAATG 468

Qy        713 AACATTGTTAGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGC 772
              |||||||  ||| |||   | ||  |  | ||      | | ||     |  ||||
Db        469 AACATTGCCAGTGCTACAATTGCACTAGTGGGGACTGCTTTTCTCTCACTAAATATAGCA 528

Qy        773 ATCAAT 778
                | |||
Db        529 GTTAAT 534


RESULT 9
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14

  Query Match             3.9%;  Score 51.6;  DB 1;  Length 7218;
  Best Local Similarity   3.1%;  Pred. No. 0.00011;
  Matches   12;  Conservative  221;  Mismatches  155;  Indels    0;  Gaps    0;

Qy        115 AGACAAGTTCTTGACTCTATGTTGAGGCCAGTTGAAAAATGAGGGAGAATAAAACCATGA 174
              | | |||   |||   | : :  ::::  ::  :::::: :::::::: :::::  : ::
Db       1451 ATAGAAGAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1392

Qy        175 ACGAAACAAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATCCAA 234
              : :::: :::::::::: :::: :::::::::: :::::::: : :: :: : :  :   ::
Db       1391 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1332

Qy        235 GCCAACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAGCTT 294
              :  :: :::  : :  :::: ::: ::::  : :   ::      :   :  ::: ::
Db       1331 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1272

Qy        295 TATGGCTCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCAGGG 354
              : ::     :::    :: ::     :::   ::  ::   :::::  :::   ::::
Db       1271 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1212

Qy        355 TGCTCAGCGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCAACC 414
              :   :: : :   ::    : :: :  : :     :::::     :  :: ::  ::
Db       1211 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1152

Qy        415 GGGTCAAGGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTTTAA 474
              :::  ::::::: : : ::: :: :::   ::: : ::::: :: :: :: :::    ::
Db       1151 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1092

Qy        475 AGAAGAAGCAAAGGCACTAGGGGTGATC 502
              :::::::: ::::: :  ::::: :|||
Db       1091 RRRRRRRRRRRRRRRRRRRRRRRRATC 1064


RESULT 10
US-09-724-864-15
; Sequence 15, Application US/09724864
; Patent No. 6380362
; GENERAL INFORMATION:
;  APPLICANT:  Watson, James D
;  APPLICANT:  Murison, James G.
;  TITLE OF INVENTION:  Polynucleotides, polypeptides expressed
;  TITLE OF INVENTION:  by the polynucleotides and methods for their use.
;  FILE REFERENCE:  11000.1050U1
;  CURRENT APPLICATION NUMBER: US/09/724,864
;  CURRENT FILING DATE:  2000-11-28
;  PRIOR APPLICATION NUMBER:  U.S. No. 6380362 60/171,678
;  PRIOR FILING DATE:  1999-12-23
;  NUMBER OF SEQ ID NOS:  72
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 15
;   LENGTH: 1308
;   TYPE: DNA
;   ORGANISM: Mouse
US-09-724-864-15

  Query Match             3.8%;  Score 49.4;  DB 4;  Length 1308;
  Best Local Similarity   50.2%;  Pred. No. 0.00018;
  Matches  150;  Conservative    0;  Mismatches  146;  Indels    3;  Gaps    1;

Qy        471 TTAAAGAAGAAGCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTG 530
              | || |  ||     || |   | | ||  ||||||||||||  ||   |||||    |
Db        186 TAAAGGCTGAGATCAAAGTGATGGCGGCAATCCAGATCATGTGTGCTGTGATGGTGTTGA 245

Qy        531 GTTTTGGAATTGTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCT 590
              || | |||||  |||||   |   |  ||| |  | |          | ||  |||   |
Db        246 GTCTGGGAATCATTTTGGCATCTGTTCCCTCCAATCTACACTTTACCTCAGTGTTT---T 302

Qy        591 CTACTGCTGTTATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCT 650
              | |      |      ||| ||||||||   || |  | | ||||  ||||||||
Db        303 CCATCCTGTTAGAATCTGGCTACCCATTTGTAGGAGCTTTGTTTTTTGCCATCTCTGGAA 362

Qy        651 CTCTCTCTGTGTCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAA 710
              |||  ||| |     ||   ||  || | |      | ||||  |  |||||| |
Db        363 TTCTGTCTATTGTCACAGAGAAAAAGATGACTAAGCCTTTGGTTCACAGCAGCCTAGCCC 422

Qy        711 TGAACATTGTTAGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATG 769
              |||  |||  | |||    |||  ||  | |  ||  |   | | ||    ||   | ||
Db        423 TGAGCATCCTGAGTGTCCTCTCTGCTCTTACAGGCATCGCTATTCTCTCTGTCAGTTTG 481


RESULT 11
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14

  Query Match             3.4%;  Score 44.4;  DB 1;  Length 7218;
  Best Local Similarity   6.4%;  Pred. No. 0.013;
  Matches   27;  Conservative  212;  Mismatches  183;  Indels    0;  Gaps    0;

Qy        511 GGTTGGATTGATGCACATTGGTTTTGGAATTGTTTTGTGTTTAATATCCTTCTCTTTTAG 570
              | |||| |  | |   |  |   |||  ||: :::: : :::  : ::::::::::::
Db       1037 GCTTGGCTGCAGGTCGAGGGAGCTTGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1096

Qy        571 AGAAGTATTAGGTTTTGCCTCTACTGCTGTTATTGGTGGATACCCATTCTGGGGTGGCCT 630
                : ::   :::: ::::: :: :: :: ::  :   : ::: ::::    :  :::
Db       1097 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1156

Qy        631 TTCTTTTATTATCTCTGGCTCTCTCTCTGTGTCAGCATCCAAGGAGCTTTCCCGTTGTCT 690
              ::::::: :: ::::: :::::::::: : ::  : :::      ::::::: :: :::
Db       1157 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1216

Qy        691 GGTGAAAGGCAGCCTGGGAATGAACATTGTTAGTTCTATCTTGGCCTTCATTGGAGTGAT 750
                :      :  :::      :   : :: ::  :::: ::::  ::::: ::    : :
Db       1217 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1276

Qy        751 TCTGCTGCTGGTGGATATGTGCATCAATGGGGTAGCTGGCCAAGACTACTGGGCCGTGCT 810
              ::: :: ::  :   : : : : ::  :    :  ::  ::    :: ::   :: : ::
Db       1277 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1336

Qy        811 TTCTGGAAAAGGCATTTCAGCCACGCTGATGATCTTCTCCCTCTTGGAGTTCTTCGTAGC 870
              ::::        : ::::  :: : ::  :  :::::::::::::    :::::: :  :
Db       1337 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1396

Qy        871 TTGTGCCACAGCCCATTTTGCCAACCAAGCAAACACCACAACCAATATGTCTGTCCTGGT 930
              :: : :: :  ::: :::: ::  ::   :   : ::   |||||  | |     ||  |
Db       1397 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYGTACCAAATTCTTCTATCTCTT 1456

Qy        931 TA 932
              ||
Db       1457 TA 1458


RESULT 12
US-08-553-619B-8
; Sequence 8, Application US/08553619B
; Patent No. 5919705
;  GENERAL INFORMATION:
;    APPLICANT:  DeHaan, Petrus T.
;    TITLE OF INVENTION:  Virus Resistant Plants
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5919705artis Crop Protection
;      STREET:  975 California Avenue
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/553,619B
;      FILING DATE:  December 1, 1995
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Marcus-Wyner, Lynn
;      REGISTRATION NUMBER:  34,869
;      REFERENCE/DOCKET NUMBER:  137-1082/PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/354-3588
;      TELEFAX:  415/857-1125
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2621 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
;    ORIGINAL SOURCE:
;      ORGANISM:  Chimeric tomato spotted wilt virus S RNA
US-08-553-619B-8

  Query Match             3.1%;  Score 40.8;  DB 2;  Length 2621;
  Best Local Similarity   51.1%;  Pred. No. 0.078;
  Matches   96;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy        112 TGAAGACAAGTTCTTGACTCTATGTTGAGGCCAGTTGAAAAATGAGGGAGAATAAAACCA 171
              | |  | |  ||  |  || | ||||| |   | |  | ||   |   | || || | |
Db        741 TAATTAAAGTTTGCTTTCTGTGTGTTGTGCTTAATAAATAATAAAATAACAAAAACAACG 800

Qy        172 TGAACGAAACAAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATC 231
                ||| ||| |  ||  ||| ||||   ||    ||  ||||  | ||||  ||   ||
Db        801 AAAACAAAAAATAAATAAAATAAAAATAAAATAAAAATAAAATAAAATAAAAATAAAATA 860

Qy        232 CAAGCCAACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAG 291
               ||   || || | |     ||  ||   |||| | ||||||   |  || |   |
Db        861 AAAAATAAAAAACAAAAAACAAAAAACAAAAACAAAACCCAAATTTGGCCAAATTGGTCC 920

Qy        292 CTTTATGG 299
              ||||  ||
Db        921 CTTTCGGG 928


RESULT 13
US-09-370-473-5
; Sequence 5, Application US/09370473
; Patent No. 6271031
; GENERAL INFORMATION:
;  APPLICANT: Falco, S. Carl
;  APPLICANT: Famodu, Layo O.
;  APPLICANT: Rafalski, J. Antoni
;  TITLE OF INVENTION: Quinolinate Metabolism Enzymes
;  FILE REFERENCE: BB-1209
;  CURRENT APPLICATION NUMBER: US/09/370,473
;  CURRENT FILING DATE: 1999-08-09
;  EARLIER APPLICATION NUMBER: 60/096,240
;  EARLIER FILING DATE: August 12, 1998
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 5
;   LENGTH: 2065
;   TYPE: DNA
;   ORGANISM: Triticum aestivum
US-09-370-473-5

  Query Match             3.1%;  Score 40.2;  DB 4;  Length 2065;
  Best Local Similarity   56.4%;  Pred. No. 0.1;
  Matches   75;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy         88 GTTCTGTTCTCTAGATTGGAGGGATGAAGACAAGTTCTTGACTCTATGTTGAGGCCAGTT 147
              ||||||  | | |    || |  |   |  | | | | |  ||| |||    | ||   ||
Db       1742 GTTCTGATATTTCAGATGTATAGGACAGCAGATGATTTGTACTTTATTCCTATGCTCATT 1801

Qy        148 GAAAAATGAGGGAGAATAAAACCATGAACGAAACAAGAAAGAAACAAAACAGAAGAGGAA 207
               |||||  |   | || ||||  |  ||  ||| || ||| ||| |||| | || |  ||
Db       1802 TAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1861

Qy        208 TGAAAAAGACATA 220
              |||||  | | |
Db       1862 AAAAAAAAAAAAA 1874
```

```
RESULT 14
US-09-797-906-1
; Sequence 1, Application US/09797906
; Patent No. 6329188
; GENERAL INFORMATION:
;  APPLICANT:  Zianghe YAN, Karen A. KETCHUM, Valentina DIFRANCESCO, Ellen M. BEASLEY
;  TITLE OF INVENTION:  ISOLATED HUMAN PROTEASE PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
;  TITLE OF INVENTION:  USES THEREOF
;  FILE REFERENCE:  CL001151CIP
;  CURRENT APPLICATION NUMBER: US/09/797,906
;  CURRENT FILING DATE:  2001-03-05
;  NUMBER OF SEQ ID NOS:  5
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 1798
;   TYPE: DNA
;   ORGANISM: Human
US-09-797-906-1

  Query Match             2.9%;  Score 38.6;  DB 4;  Length 1798;
  Best Local Similarity   54.6%;  Pred. No. 0.27;
  Matches   77;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy        133 ATGTTGAGGCCAGTTGAAAAATGAGGGAGAATAAAACCATGAACGAAACAAGAAAGAAAC 192
              ||||    | |||| ||||||  |   | || ||||  |  ||  ||| || ||| |||
Db       1629 ATGTCCTTGGCAGTGGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1688

Qy        193 AAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATCCAAGCCAACAAGCCATGCTGA 252
              |||| | || |  ||  ||||| | | ||  |    |   ||   || ||   |     |
Db       1689 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1748

Qy        253 AGTAAATGAAACCATACCCAA 273
              |  |||  |||  | |   ||
Db       1749 AAAAAAAAAAAAAAAAAAAA 1769


RESULT 15
5221620-3
;Patent No. 5221620
;    APPLICANT: PURCHIO,ANTHONY F.;MADISEN, LINDA;WEBB, NANCY
;    TITLE OF INVENTION: CLONING AND EXPRESSION OF TRANSFORMING
;GROWTH FACTOR BETA-2
;    NUMBER OF SEQUENCES: 16
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/446,020
;      FILING DATE: 05-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 285,140
;      FILING DATE: 16-DEC-1988
;      APPLICATION NUMBER: 234,065
;      FILING DATE: 18-AUG-1988
;      APPLICATION NUMBER: 148,267
;      FILING DATE: 25-JAN-1988
;      APPLICATION NUMBER: 106,752
;      FILING DATE: 06-OCT-1987
;SEQ ID NO:3:
;      LENGTH: 2207
5221620-3

  Query Match             2.9%;  Score 38.4;  DB 6;  Length 2207;
  Best Local Similarity   50.5%;  Pred. No. 0.35;
  Matches   93;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy         59 CACAAACTGTGGCTGGTCTACCTGTGTGGGTTCTGTTCTCTAGATTGGAGGGATGAAGAC 118
              || | ||| |||  | |     | |||  |||         |  |   |   |  |||
Db       1579 CAGACACTTTGGAAGTTTGTGTTCTGTTTGTTAAAACTGGCATCTGACACAAAAAAAGTT 1638

Qy        119 AAGTTCTTGACTCTATGTTGAGGCCAGTTGAAAAATGAGGGAGAATAAAACCATGAACGA 178
               |   | | | ||||  ||    | | ||   || |||| ||||  | || ||
Db       1639 GAAGGCCTTATTCTACATTTCACCTACTTTGTAAGTGAGAGAGACAAGAAGCAAATTTTT 1698

Qy        179 AACAAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATCCAAGCCA 238
              ||  || |||  |||||   || ||||  |  ||   |||| |||| | | | | ||
Db       1699 TTTAAAGAAAAAAATAAACACTGGAAGAATTTATTAGTGTTAATTATGTGAACAACGACA 1758

Qy        239 ACAA 242
              ||||
Db       1759 ACAA 1762


Search completed: July 16, 2003, 14:58:50
Job time : 53.3795 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 16, 2003, 13:28:07 ; Search time 173.297 Seconds
(without alignments)
12133.023 Million cell updates/sec

Title: US-09-981-353-81
Perfect score: 1317
Sequence: 1 aaggaacaaagtaagtccat.....tttttccttaataaaatgtc 1317

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1114047 seqs, 798260406 residues

Total number of hits satisfying chosen parameters: 2228094

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
6: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1317	100.0	1317	9	US-09-981-353-81	Sequence 81, Appl
2	1317	100.0	1318	9	US-10-158-646-77	Sequence 77, Appl
3	1084.4	82.3	1291	9	US-10-054-988-17	Sequence 17, Appl
4	1084.4	82.3	1291	10	US-09-739-254-17	Sequence 17, Appl
5	1084.4	82.3	1291	10	US-09-904-615-17	Sequence 17, Appl
6	1079	81.9	1292	9	US-10-054-988-61	Sequence 61, Appl
7	1079	81.9	1292	10	US-09-739-254-61	Sequence 61, Appl
8	1079	81.9	1292	10	US-09-904-615-61	Sequence 61, Appl
9	389.6	29.6	441	9	US-09-925-299-727	Sequence 727, App
10	389.6	29.6	441	10	US-09-925-299-727	Sequence 727, App
11	303.6	23.1	351	9	US-09-803-719-494	Sequence 494, App
12	292.6	22.2	342	9	US-09-803-719-638	Sequence 638, App
13	288.6	21.9	383	9	US-09-803-719-481	Sequence 481, App
14	92.6	7.0	125	10	US-09-563-817-780	Sequence 780, App
15	81.2	6.2	753	9	US-10-017-754-2003	Sequence 2003, Ap
16	81.2	6.2	1236	9	US-10-156-136-5	Sequence 5, Appli
c 17	81.2	6.2	1236	9	US-10-156-136-51	Sequence 51, Appl
18	81.2	6.2	1336	9	US-09-736-457-1676	Sequence 1676, Ap
19	81.2	6.2	1336	9	US-09-902-941-1676	Sequence 1676, Ap
20	81.2	6.2	1336	9	US-09-849-626-1676	Sequence 1676, Ap
21	81.2	6.2	1336	9	US-10-017-754-1676	Sequence 1676, Ap
22	81.2	6.2	1374	10	US-09-925-297-305	Sequence 305, App
23	79.6	6.0	861	9	US-09-902-941-1877	Sequence 1877, Ap
24	79.6	6.0	861	9	US-09-849-626-1877	Sequence 1877, Ap
25	79.6	6.0	861	9	US-10-017-754-1877	Sequence 1877, Ap
26	79.6	6.0	1155	9	US-09-902-941-1875	Sequence 1875, Ap
27	79.6	6.0	1155	9	US-09-849-626-1875	Sequence 1875, Ap
28	79.6	6.0	1155	9	US-10-017-754-1875	Sequence 1875, Ap
29	79.6	6.0	1353	9	US-09-902-941-1873	Sequence 1873, Ap
30	79.6	6.0	1353	9	US-09-849-626-1873	Sequence 1873, Ap
31	79.6	6.0	1353	9	US-10-017-754-1873	Sequence 1873, Ap
32	74.8	5.7	932	9	US-10-220-946-7	Sequence 7, Appli
33	74.8	5.7	945	9	US-09-809-391-168	Sequence 168, App
34	74.8	5.7	1036	9	US-10-220-946-5	Sequence 5, Appli
35	74.8	5.7	1728	10	US-09-822-849A-23	Sequence 23, Appl
36	74.8	5.7	1743	10	US-09-925-302-314	Sequence 314, App
37	74.8	5.7	1871	9	US-10-106-698-2116	Sequence 2116, Ap
c 38	60	4.6	486	9	US-10-046-935-2044	Sequence 2044, Ap
c 39	60	4.6	486	9	US-09-878-178-2044	Sequence 2044, Ap
c 40	60	4.6	486	9	US-10-146-502-2044	Sequence 2044, Ap
41	53.8	4.1	1146	9	US-10-207-655-130	Sequence 130, App
42	53	4.0	1330	9	US-09-374-046A-151	Sequence 151, App
43	53	4.0	2438	9	US-09-822-846-227	Sequence 227, App
44	52.4	4.0	747	10	US-09-731-872-17	Sequence 17, Appl
45	52.4	4.0	760	10	US-09-821-821-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-981-353-81
; Sequence 81, Application US/09981353
; Patent No. US20020160382A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy W.
; APPLICANT: Jones, David A.
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
; FILE REFERENCE: PA-0038 US
; CURRENT APPLICATION NUMBER: US/09/981,353
; CURRENT FILING DATE: 2001-10-11
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 81
; LENGTH: 1317
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020160382A1 611082CB1
US-09-981-353-81

Query Match 100.0%; Score 1317; DB 9; Length 1317;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AAGGAACAAAGTAAGTCCATTGATACGTTCTTGCCTATCTCTCCTCCAAATCAATGGGCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGGAACAAAGTAAGTCCATTGATACGTTCTTGCCTATCTCTCCTCCAAATCAATGGGCA 60

Qy         61 CAAACTGTGGCTGGTCTACCTGTGTGGGTTCTGTTCTCTAGATTGGAGGGATGAAGACAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAACTGTGGCTGGTCTACCTGTGTGGGTTCTGTTCTCTAGATTGGAGGGATGAAGACAA 120

Qy        121 GTTCTTGACTCTATGTTGAGGCCAGTTGAAAAATGAGGGAGAATAAAACCATGAACGAAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTTCTTGACTCTATGTTGAGGCCAGTTGAAAAATGAGGGAGAATAAAACCATGAACGAAA 180

Qy        181 CAAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATCCAAGCCAAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATCCAAGCCAAC 240
```

```
Qy       241 AAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAGCTTTATGGC 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 AAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAGCTTTATGGC 300

Qy       301 TCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCAGGGTGCTCA 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 TCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCAGGGTGCTCA 360

Qy       361 GCGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCAACCGGGTCA 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 GCGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCAACCGGGTCA 420

Qy       421 AGGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTTTAAAGAAGA 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 AGGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTTTAAAGAAGA 480

Qy       481 AGCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGGTTTTGGAAT 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 AGCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGGTTTTGGAAT 540

Qy       541 TGTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCTACTGCTGT 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 TGTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCTACTGCTGT 600

Qy       601 TATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCTCTCTCTGT 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 TATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCTCTCTCTGT 660

Qy       661 GTCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATGAACATTGT 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 GTCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATGAACATTGT 720

Qy       721 TAGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGCATCAATGG 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 TAGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGCATCAATGG 780

Qy       781 GGTAGCTGGCCAAGACTACTGGGCCGTGCTTTCTGGAAAAGGCATTTCAGCCACGCTGAT 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 GGTAGCTGGCCAAGACTACTGGGCCGTGCTTTCTGGAAAAGGCATTTCAGCCACGCTGAT 840

Qy       841 GATCTTCTCCCTCTTGGAGTTCTTCGTAGCTTGTGCCACAGCCCATTTTGCCAACCAAGC 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 GATCTTCTCCCTCTTGGAGTTCTTCGTAGCTTGTGCCACAGCCCATTTTGCCAACCAAGC 900

Qy       901 AAACACCACAACCAATATGTCTGTCCTGGTTATTCCAAATATGTATGAAAGCAACCCTGT 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 AAACACCACAACCAATATGTCTGTCCTGGTTATTCCAAATATGTATGAAAGCAACCCTGT 960

Qy       961 GACACCAGCGTCTTCTTCAGCTCCTCCCAGATGCAACAACTACTCAGCTAATGCCCCTAA 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 GACACCAGCGTCTTCTTCAGCTCCTCCCAGATGCAACAACTACTCAGCTAATGCCCCTAA 1020

Qy      1021 ATAGTAAAAGAAAAAGGGGTATCAGTCTAATCTCATGGAGAAAAACTACTTGCAAAAACT 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1021 ATAGTAAAAGAAAAAGGGGTATCAGTCTAATCTCATGGAGAAAAACTACTTGCAAAAACT 1080

Qy      1081 TCTTAAGAAGATGTCTTTTATTGTCTACAATGATTTCTAGTCTTTAAAAACTGTGTTTGA 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1081 TCTTAAGAAGATGTCTTTTATTGTCTACAATGATTTCTAGTCTTTAAAAACTGTGTTTGA 1140

Qy      1141 GATTTGTTTTTAGGTTGGTCGCTAATGATGGCTGTATCTCCCTTCACTGTCTCTTCCTAC 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1141 GATTTGTTTTTAGGTTGGTCGCTAATGATGGCTGTATCTCCCTTCACTGTCTCTTCCTAC 1200

Qy      1201 ATTACCACTACTACATGCTGGCAAAGGTGAAGGATCAGAGGACTGAAAAATGATTCTGCA 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1201 ATTACCACTACTACATGCTGGCAAAGGTGAAGGATCAGAGGACTGAAAAATGATTCTGCA 1260

Qy      1261 ACTCTCTTAAAGTTAGAAATGTTTCTGTTCATATTACTTTTTCCTTAATAAAATGTC 1317
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1261 ACTCTCTTAAAGTTAGAAATGTTTCTGTTCATATTACTTTTTCCTTAATAAAATGTC 1317
```

```
RESULT 2
US-10-158-646-77
; Sequence 77, Application US/10158646
; Publication No. US20030073105A1
; GENERAL INFORMATION:
;  APPLICANT: Lasek, Amy K.W.
;  APPLICANT:  Sornasse, Thierry
;  TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
;  FILE REFERENCE: PA-0030-1 US
;  CURRENT APPLICATION NUMBER: US/10/158,646
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: 60/295,239
;  PRIOR FILING DATE: 2001-05-31
;  NUMBER OF SEQ ID NOS: 78
;  SOFTWARE: PERL Program
; SEQ ID NO 77
;   LENGTH: 1318
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20030073105A1 417113.5
US-10-158-646-77

  Query Match            100.0%;  Score 1317;  DB 9;  Length 1318;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1317;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGGAACAAAGTAAGTCCATTGATACGTTCTTGCCTATCTCTCCTCCAAATCAATGGGCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 AAGGAACAAAGTAAGTCCATTGATACGTTCTTGCCTATCTCTCCTCCAAATCAATGGGCA 61

Qy         61 CAAACTGTGGCTGGTCTACCTGTGTGGGTTCTGTTCTCTAGATTGGAGGGATGAAGACAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 CAAACTGTGGCTGGTCTACCTGTGTGGGTTCTGTTCTCTAGATTGGAGGGATGAAGACAA 121

Qy        121 GTTCTTGACTCTATGTTGAGGCCAGTTGAAAAATGAGGGAGAATAAAACCATGAACGAAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 GTTCTTGACTCTATGTTGAGGCCAGTTGAAAAATGAGGGAGAATAAAACCATGAACGAAA 181

Qy        181 CAAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATCCAAGCCAAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 CAAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATCCAAGCCAAC 241

Qy        241 AAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAGCTTTATGGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 AAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAGCTTTATGGC 301

Qy        301 TCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCAGGGTGCTCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 TCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCAGGGTGCTCA 361

Qy        361 GCGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCAACCGGGTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 GCGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCAACCGGGTCA 421

Qy        421 AGGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTTTAAAGAAGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        422 AGGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTTTAAAGAAGA 481

Qy        481 AGCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGGTTTTGGAAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        482 AGCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGGTTTTGGAAT 541

Qy        541 TGTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCTACTGCTGT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        542 TGTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCTACTGCTGT 601

Qy        601 TATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCTCTCTCTGT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        602 TATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCTCTCTCTGT 661

Qy        661 GTCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATGAACATTGT 720
```

```
Db        662 GTCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATGAACATTGT 721

Qy        721 TAGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGCATCAATGG 780

Db        722 TAGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGCATCAATGG 781

Qy        781 GGTAGCTGGCCAAGACTACTGGGCCGTGCTTTCTGGAAAAGGCATTTCAGCCACGCTGAT 840

Db        782 GGTAGCTGGCCAAGACTACTGGGCCGTGCTTTCTGGAAAAGGCATTTCAGCCACGCTGAT 841

Qy        841 GATCTTCTCCCTCTTGGAGTTCTTCGTAGCTTGTGCCACAGCCCATTTTGCCAACCAAGC 900

Db        842 GATCTTCTCCCTCTTGGAGTTCTTCGTAGCTTGTGCCACAGCCCATTTTGCCAACCAAGC 901

Qy        901 AAACACCACAACCAATATGTCTGTCCTGGTTATTCCAAATATGTATGAAAGCAACCCTGT 960

Db        902 AAACACCACAACCAATATGTCTGTCCTGGTTATTCCAAATATGTATGAAAGCAACCCTGT 961

Qy        961 GACACCAGCGTCTTCTTCAGCTCCTCCCAGATGCAACAACTACTCAGCTAATGCCCCTAA 1020

Db        962 GACACCAGCGTCTTCTTCAGCTCCTCCCAGATGCAACAACTACTCAGCTAATGCCCCTAA 1021

Qy       1021 ATAGTAAAAGAAAAAGGGGTATCAGTCTAATCTCATGGAGAAAAACTACTTGCAAAAACT 1080

Db       1022 ATAGTAAAAGAAAAAGGGGTATCAGTCTAATCTCATGGAGAAAAACTACTTGCAAAAACT 1081

Qy       1081 TCTTAAGAAGATGTCTTTTATTGTCTACAATGATTTCTAGTCTTTAAAAACTGTGTTTGA 1140

Db       1082 TCTTAAGAAGATGTCTTTTATTGTCTACAATGATTTCTAGTCTTTAAAAACTGTGTTTGA 1141

Qy       1141 GATTTGTTTTTAGGTTGGTCGCTAATGATGGCTGTATCTCCCTTCACTGTCTCTTCCTAC 1200

Db       1142 GATTTGTTTTTAGGTTGGTCGCTAATGATGGCTGTATCTCCCTTCACTGTCTCTTCCTAC 1201

Qy       1201 ATTACCACTACTACATGCTGGCAAAGGTGAAGGATCAGAGGACTGAAAAATGATTCTGCA 1260

Db       1202 ATTACCACTACTACATGCTGGCAAAGGTGAAGGATCAGAGGACTGAAAAATGATTCTGCA 1261

Qy       1261 ACTCTCTTAAAGTTAGAAATGTTTCTGTTCATATTACTTTTTCCTTAATAAAATGTC 1317

Db       1262 ACTCTCTTAAAGTTAGAAATGTTTCTGTTCATATTACTTTTTCCTTAATAAAATGTC 1318
```

```
RESULT 3
US-10-054-988-17
; Sequence 17, Application US/10054988
; Publication No. US20030087341A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: 49 Human Secreted Proteins
;  FILE REFERENCE: PZ032P1
;  CURRENT APPLICATION NUMBER: US/10/054,988
;  CURRENT FILING DATE:  2002-01-25
;  PRIOR APPLICATION NUMBER: 09/904,615
;  PRIOR FILING DATE: 2001-07-16
;  PRIOR APPLICATION NUMBER: 09/511,554
;  PRIOR FILING DATE: 2000-02-23
;  PRIOR APPLICATION NUMBER: 60/097,917
;  PRIOR FILING DATE: 1998-08-25
;  PRIOR APPLICATION NUMBER: 60/098,634
;  PRIOR FILING DATE: 1998-08-31
;  NUMBER OF SEQ ID NOS: 170
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
;   LENGTH: 1291
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (1279)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (1286)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (1290)
;   OTHER INFORMATION: n equals a,t,g, or c
US-10-054-988-17

  Query Match             82.3%;  Score 1084.4;  DB 9;  Length 1291;
  Best Local Similarity   97.8%;  Pred. No. 6.2e-297;
  Matches 1111;  Conservative    1;  Mismatches   17;  Indels    7;  Gaps    1;

Qy        182 AAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATCCAAGCCAACA 241

Db         75 AAAGAGGAAACATAGAGGTGCCAAAGGAACAAAGACATAATGATGTCATCCAAGCCAACA 134

Qy        242 AGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAGCTTTATGGCT 301

Db        135 AGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAGCTTTATGGCT 194

Qy        302 CCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCAGGGTGCTCAG 361

Db        195 CCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCAGGGTGCTCAG 254

Qy        362 CGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCAACCGGGTCAA 421

Db        255 CGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCAACCGGGTCAA 314

Qy        422 GGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTTTAAAGAAGAA 481

Db        315 GGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTTTAAAGAAGAA 374

Qy        482 GCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGGTTTTGGAATT 541

Db        375 GCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGGTTTTGGAATT 434

Qy        542 GTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCTACTGCTGTT 601

Db        435 GTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCTACTGCTGTT 494

Qy        602 ATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCTCTCTCTGTG 661

Db        495 ATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCTCTCTCTGTG 554

Qy        662 TCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATGAACATTGTT 721

Db        555 TCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATGAACATTGKT 614

Qy        722 AGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGCATCAATGGG 781

Db        615 AGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGCATCAATGGG 674

Qy        782 GTAGCTGGCCAAGACTACTGGGCCGTGCTTTCTGGAAAAGGCATTTCAGCCACGCTGATG 841

Db        675 GTAGCTGGCCAAGACTACTGGGCCGTGCTTTCTGGAAAAGGCATTTCAGCCACGCTGATG 734

Qy        842 ATCTTCTCCCTCTTGGAGTTCTTCGTAGCTTGTGCCACAGCCCATTTTGCCAACCAAGCA 901

Db        735 ATCTTCTCCCTCTTGGAGTTCTTCGTAGCTTGTGCCACAGCCCATTTTGCCAACCAAGCA 794

Qy        902 AACACCACAACCAATATGTCTGTCCTGGTTATTCCAAATATGTATGAAAGCAACCCTGTG 961

Db        795 AACACCACAACCAATATGTCTGTCCTGGTTATTCCAAATATGTATGAAAGCAACCCTGTG 854

Qy        962 ACACCAGCGTCTTCTTCAGCTCCTCCCAGATGCAACAACTACTCAGCTAATGCCCCTAAA 1021

Db        855 ACACCAGCGTCTTCTTCAGCTCCTCCCAGATGCAACAACTACTCAGCTAATGCCCC---- 910

Qy       1022 TAGTAAAAGAAAAAGGGGTATCAGTCTAATCTCATGGAGAAAAACTACTTGCAAAAACTT 1081

Db        911 ---TAAAAGAAAAAGGGGTATCAGTCTAATCTCATGGAGAAAAACTACTTGCAAAAACTT 967

Qy       1082 CTTAAGAAGATGTCTTTTATTGTCTACAATGATTTCTAGTCTTTAAAAACTGTGTTTGAG 1141
```

```
Db        968 CTTAAGAAGATGTCTTTTATTGTCTACAATGATTTCTAGTCTTTAAAAACTGTGTTTGAG 1027

Qy       1142 ATTTGTTTTTAGGTTGGTCGCTAATGATGGCTGTATCTCCCTTCACTGTCTCTTCCTACA 1201
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1028 ATTTGTTTTTAGGTTGGTCGCTAATGATGGCTGTATCTCCCTTCACTGTCTCTTCCTACA 1087

Qy       1202 TTACCACTACTACATGCTGGCAAAGGTGAAGGATCAGAGGACTGAAAAATGATTCTGCAA 1261
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1088 TTACCACTACTACATGCTGGCAAAGGTGAAGGATCAGAGGACTGAAAAATGATTCTGCAA 1147

Qy       1262 CTCTCTTAAAGTTAGAAATGTTTCTGTTCATATTACTTTTTCCTTAATAAAATGTC 1317
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1148 CTCTCTTAAAGTTAGAAATGTTTCTGTTCATATTACTTTTTCCTTAATAAAATGTC 1203


RESULT 4
US-09-739-254-17
; Sequence 17, Application US/09739254
; Patent No. US20010021700A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: 49 Human Secreted Proteins
;  FILE REFERENCE: PZ032P1
;  CURRENT APPLICATION NUMBER: US/09/739,254
;  CURRENT FILING DATE: 2000-12-19
;  EARLIER APPLICATION NUMBER: 09/511,554
;  EARLIER FILING DATE: 2000-02-23
;  EARLIER APPLICATION NUMBER: PCT/US99/19330
;  EARLIER FILING DATE: 1999-08-24
;  EARLIER APPLICATION NUMBER: 60/097,917
;  EARLIER FILING DATE: 1998-08-25
;  EARLIER APPLICATION NUMBER: 60/098,634
;  EARLIER FILING DATE: 1998-08-31
;  NUMBER OF SEQ ID NOS: 170
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
;   LENGTH: 1291
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (1279)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (1286)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (1290)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-739-254-17

  Query Match             82.3%;  Score 1084.4;  DB 10;  Length 1291;
  Best Local Similarity   97.8%;  Pred. No. 6.2e-297;
  Matches 1111;  Conservative    1;  Mismatches   17;  Indels    7;  Gaps    1;

Qy        182 AAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATCCAAGCCAACA 241
              ||  | ||||| |   |      |   | ||||||||||||||||||||||||||||||||
Db         75 AAAGAGGAAACATAGAGGTGCCAAAGGAACAAAGACATAATGATGTCATCCAAGCCAACA 134

Qy        242 AGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAGCTTTATGGCT 301
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        135 AGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAGCTTTATGGCT 194

Qy        302 CCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCAGGGTGCTCAG 361
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        195 CCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCAGGGTGCTCAG 254

Qy        362 CGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCAACCGGGTCAA 421
              . ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        255 CGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCAACCGGGTCAA 314

Qy        422 GGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTTTAAAGAAGAA 481
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        315 GGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTTTAAAGAAGAA 374

Qy        482 GCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGGTTTTGGAATT 541
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        375 GCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGGTTTTGGAATT 434

Qy        542 GTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCTACTGCTGTT 601
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        435 GTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCTACTGCTGTT 494

Qy        602 ATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCTCTCTCTGTG 661
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        495 ATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCTCTCTCTGTG 554

Qy        662 TCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATGAACATTGTT 721
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db        555 TCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATGAACATTGKT 614

Qy        722 AGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGCATCAATGGG 781
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        615 AGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGCATCAATGGG 674

Qy        782 GTAGCTGGCCAAGACTACTGGGCCGTGCTTTCTGGAAAAGGCATTTCAGCCACGCTGATG 841
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        675 GTAGCTGGCCAAGACTACTGGGCCGTGCTTTCTGGAAAAGGCATTTCAGCCACGCTGATG 734

Qy        842 ATCTTCTCCCTCTTGGAGTTCTTCGTAGCTTGTGCCACAGCCCATTTTGCCAACCAAGCA 901
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        735 ATCTTCTCCCTCTTGGAGTTCTTCGTAGCTTGTGCCACAGCCCATTTTGCCAACCAAGCA 794

Qy        902 AACACCACAACCAATATGTCTGTCCTGGTTATTCCAAATATGTATGAAAGCAACCCTGTG 961
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        795 AACACCACAACCAATATGTCTGTCCTGGTTATTCCAAATATGTATGAAAGCAACCCTGTG 854

Qy        962 ACACCAGCGTCTTCTTCAGCTCCTCCCAGATGCAACAACTACTCAGCTAATGCCCCTAAA 1021
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        855 ACACCAGCGTCTTCTTCAGCTCCTCCCAGATGCAACAACTACTCAGCTAATGCCCC---- 910

Qy       1022 TAGTAAAAGAAAAAGGGGTATCAGTCTAATCTCATGGAGAAAAACTACTTGCAAAAACTT 1081
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        911 ---TAAAAGAAAAAGGGGTATCAGTCTAATCTCATGGAGAAAAACTACTTGCAAAAACTT 967

Qy       1082 CTTAAGAAGATGTCTTTTATTGTCTACAATGATTTCTAGTCTTTAAAAACTGTGTTTGAG 1141
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        968 CTTAAGAAGATGTCTTTTATTGTCTACAATGATTTCTAGTCTTTAAAAACTGTGTTTGAG 1027

Qy       1142 ATTTGTTTTTAGGTTGGTCGCTAATGATGGCTGTATCTCCCTTCACTGTCTCTTCCTACA 1201
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1028 ATTTGTTTTTAGGTTGGTCGCTAATGATGGCTGTATCTCCCTTCACTGTCTCTTCCTACA 1087

Qy       1202 TTACCACTACTACATGCTGGCAAAGGTGAAGGATCAGAGGACTGAAAAATGATTCTGCAA 1261
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1088 TTACCACTACTACATGCTGGCAAAGGTGAAGGATCAGAGGACTGAAAAATGATTCTGCAA 1147

Qy       1262 CTCTCTTAAAGTTAGAAATGTTTCTGTTCATATTACTTTTTCCTTAATAAAATGTC 1317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1148 CTCTCTTAAAGTTAGAAATGTTTCTGTTCATATTACTTTTTCCTTAATAAAATGTC 1203


RESULT 5
US-09-904-615-17
; Sequence 17, Application US/09904615
; Patent No. US20020026040A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: 49 Human Secreted Proteins
;  FILE REFERENCE: PZ032P1
;  CURRENT APPLICATION NUMBER: US/09/904,615
;  CURRENT FILING DATE:  2001-07-16
;  PRIOR APPLICATION NUMBER: 09/511,554
;  PRIOR FILING DATE: 2000-02-23
```

```
;  PRIOR APPLICATION NUMBER: 60/097,917
;  PRIOR FILING DATE: 1998-08-25
;  PRIOR APPLICATION NUMBER: 60/098,634
;  PRIOR FILING DATE: 1998-08-31
;  NUMBER OF SEQ ID NOS: 170
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
;   LENGTH: 1291
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (1279)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (1286)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (1290)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-904-615-17

  Query Match             82.3%;  Score 1084.4;  DB 10;  Length 1291;
  Best Local Similarity   97.8%;  Pred. No. 6.2e-297;
  Matches 1111;  Conservative    1;  Mismatches   17;  Indels    7;  Gaps    1;

Qy        182 AAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATCCAAGCCAACA 241
              ||  | ||||| |   |      |  | ||||||||||||||||||||||||||||||
Db         75 AAAGAGGAAACATAGAGGTGCCAAAGGAACAAAGACATAATGATGTCATCCAAGCCAACA 134

Qy        242 AGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAGCTTTATGGCT 301
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        135 AGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAGCTTTATGGCT 194

Qy        302 CCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCAGGGTGCTCAG 361
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        195 CCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCAGGGTGCTCAG 254

Qy        362 CGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCAACCGGGTCAA 421
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        255 CGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCAACCGGGTCAA 314

Qy        422 GGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTTTAAAGAAGAA 481
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        315 GGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTTTAAAGAAGAA 374

Qy        482 GCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGGTTTTGGAATT 541
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        375 GCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGGTTTTGGAATT 434

Qy        542 GTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCTACTGCTGTT 601
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        435 GTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCTACTGCTGTT 494

Qy        602 ATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCTCTCTCTGTG 661
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        495 ATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCTCTCTCTGTG 554

Qy        662 TCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATGAACATTGTT 721
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db        555 TCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATGAACATTGKT 614

Qy        722 AGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGCATCAATGGG 781
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        615 AGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGCATCAATGGG 674

Qy        782 GTAGCTGGCCAAGACTACTGGGCCGTGCTTTCTGGAAAAGGCATTTCAGCCACGCTGATG 841
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        675 GTAGCTGGCCAAGACTACTGGGCCGTGCTTTCTGGAAAAGGCATTTCAGCCACGCTGATG 734

Qy        842 ATCTTCTCCCTCTTGGAGTTCTTCGTAGCTTGTGCCACAGCCCATTTTGCCAACCAAGCA 901
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        735 ATCTTCTCCCTCTTGGAGTTCTTCGTAGCTTGTGCCACAGCCCATTTTGCCAACCAAGCA 794

Qy        902 AACACCACAACCAATATGTCTGTCCTGGTTATTCCAAATATGTATGAAAGCAACCCTGTG 961
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        795 AACACCACAACCAATATGTCTGTCCTGGTTATTCCAAATATGTATGAAAGCAACCCTGTG 854

Qy        962 ACACCAGCGTCTTCTTCAGCTCCTCCCAGATGCAACAACTACTCAGCTAATGCCCCTAAA 1021
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        855 ACACCAGCGTCTTCTTCAGCTCCTCCCAGATGCAACAACTACTCAGCTAATGCCCC---- 910

Qy       1022 TAGTAAAAGAAAAAGGGGTATCAGTCTAATCTCATGGAGAAAAACTACTTGCAAAAACTT 1081
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        911 ---TAAAAGAAAAAGGGGTATCAGTCTAATCTCATGGAGAAAAACTACTTGCAAAAACTT 967

Qy       1082 CTTAAGAAGATGTCTTTTATTGTCTACAATGATTTCTAGTCTTTAAAAACTGTGTTTGAG 1141
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        968 CTTAAGAAGATGTCTTTTATTGTCTACAATGATTTCTAGTCTTTAAAAACTGTGTTTGAG 1027

Qy       1142 ATTTGTTTTTAGGTTGGTCGCTAATGATGGCTGTATCTCCCTTCACTGTCTCTTCCTACA 1201
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1028 ATTTGTTTTTAGGTTGGTCGCTAATGATGGCTGTATCTCCCTTCACTGTCTCTTCCTACA 1087

Qy       1202 TTACCACTACTACATGCTGGCAAAGGTGAAGGATCAGAGGACTGAAAAATGATTCTGCAA 1261
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1088 TTACCACTACTACATGCTGGCAAAGGTGAAGGATCAGAGGACTGAAAAATGATTCTGCAA 1147

Qy       1262 CTCTCTTAAAGTTAGAAATGTTTCTGTTCATATTACTTTTTCCTTAATAAAATGTC 1317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1148 CTCTCTTAAAGTTAGAAATGTTTCTGTTCATATTACTTTTTCCTTAATAAAATGTC 1203


RESULT 6
US-10-054-988-61
; Sequence 61, Application US/10054988
; Publication No. US20030087341A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: 49 Human Secreted Proteins
;  FILE REFERENCE: PZ032P1
;  CURRENT APPLICATION NUMBER: US/10/054,988
;  CURRENT FILING DATE:  2002-01-25
;  PRIOR APPLICATION NUMBER: 09/904,615
;  PRIOR FILING DATE: 2001-07-16
;  PRIOR APPLICATION NUMBER: 09/511,554
;  PRIOR FILING DATE: 2000-02-23
;  PRIOR APPLICATION NUMBER: 60/097,917
;  PRIOR FILING DATE: 1998-08-25
;  PRIOR APPLICATION NUMBER: 60/098,634
;  PRIOR FILING DATE: 1998-08-31
;  NUMBER OF SEQ ID NOS: 170
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 61
;   LENGTH: 1292
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (71)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (697)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (1280)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (1287)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (1291)
;   OTHER INFORMATION: n equals a,t,g, or c
US-10-054-988-61

  Query Match             81.9%;  Score 1079;  DB 9;  Length 1292;
```

Best Local Similarity 97.3%; Pred. No. 2.1e-295;
Matches 1105; Conservative 4; Mismatches 20; Indels 7; Gaps 1;

```
Qy       182 AAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATCCAAGCCAACA 241
             ||  | |||||| |   |      |   | ||||||||||||||||||||||||||||||
Db        76 AAAGAGGAAACATAGAGGTGCCAAAGGAACAAAGACATAATGATGTCATCCAAGCCAACA 135

Qy       242 AGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAGCTTTATGGCT 301
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       136 AGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAGCTTTATGGCT 195

Qy       302 CCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCAGGGTGCTCAG 361
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       196 CCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCAGGGTGCTCAG 255

Qy       362 CGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCAACCGGGTCAA 421
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       256 CGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCAACCGGGTCAA 315

Qy       422 GGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTTTAAAGAAGAA 481
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       316 GGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTTTAAAGAAGAA 375

Qy       482 GCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGGTTTTGGAATT 541
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       376 GCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGGTTTTGGAATT 435

Qy       542 GTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCTACTGCTGTT 601
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db       436 GTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCTACTGCTGKT 495

Qy       602 ATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCTCTCTCTGTG 661
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       496 ATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCTCTCTCTGTG 555

Qy       662 TCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATGAACATTGTT 721
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db       556 TCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATGAACATTGGT 615

Qy       722 AGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGCATCAATGGG 781
             || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       616 AGGTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGCATCAATGGG 675

Qy       782 GTAGCTGGCCAAGACTACTGGGCCGTGCTTTCTGGAAAAGGCATTTCAGCCACGCTGATG 841
             |||:|||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       676 GTARCTGGCCAAGACTACTGGNCCGTGCTTTCTGGAAAAGGCATTTCAGCCACGCTGATG 735

Qy       842 ATCTTCTCCCTCTTGGAGTTCTTCGTAGCTTGTGCCACAGCCCATTTTGCCAACCAAGCA 901
             ||||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||
Db       736 ATCTTCTCCYTCTTGGAGTTCTTCGTAGCTTGTGCCACAGCCCATTTTGCCAACCAAGCA 795

Qy       902 AACACCACAACCAATATGTCTGTCCTGGTTATTCCAAATATGTATGAAAGCAACCCTGTG 961
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db       796 AACACCACAACCAATATGTCTGTCCTGGTTATTCCAAATATGTATGAAAGCAACCCTGKG 855

Qy       962 ACACCAGCGTCTTCTTCAGCTCCTCCCAGATGCAACAACTACTCAGCTAATGCCCCTAAA 1021
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       856 ACACCAGCGTCTTCTTCAGCTCCTCCCAGATGCAACAACTACTCAGCTAATGCCCC---- 911

Qy      1022 TAGTAAAAGAAAAAGGGGTATCAGTCTAATCTCATGGAGAAAAACTACTTGCAAAAACTT 1081
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       912 ---TAAAAGAAAAAGGGGTATCAGTCTAATCTCATGGAGAAAAACTACTTGCAAAAACTT 968

Qy      1082 CTTAAGAAGATGTCTTTTATTGTCTACAATGATTTCTAGTCTTTAAAAACTGTGTTTGAG 1141
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       969 CTTAAGAAGATGTCTTTTATTGTCTACAATGATTTCTAGTCTTTAAAAACTGTGTTTGAG 1028

Qy      1142 ATTTGTTTTTAGGTTGGTCGCTAATGATGGCTGTATCTCCCTTCACTGTCTCTTCCTACA 1201
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1029 ATTTGTTTTTAGGTTGGTCGCTAATGATGGCTGTATCTCCCTTCACTGTCTCTTCCTACA 1088

Qy      1202 TTACCACTACTACATGCTGGCAAAGGTGAAGGATCAGAGGACTGAAAAATGATTCTGCAA 1261
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1089 TTACCACTACTACATGCTGGCAAAGGTGAAGGATCAGAGGACTGAAAAATGATTCTGCAA 1148

Qy      1262 CTCTCTTAAAGTTAGAAATGTTTCTGTTCATATTACTTTTTCCTTAATAAAATGTC 1317
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1149 CTCTCTTAAAGTTAGAAATGTTTCTGTTCATATTACTTTTTCCTTAATAAAATGTC 1204
```

RESULT 7
US-09-739-254-61
; Sequence 61, Application US/09739254
; Patent No. US20010021700A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 49 Human Secreted Proteins
; FILE REFERENCE: PZ032P1
; CURRENT APPLICATION NUMBER: US/09/739,254
; CURRENT FILING DATE: 2000-12-19
; EARLIER APPLICATION NUMBER: 09/511,554
; EARLIER FILING DATE: 2000-02-23
; EARLIER APPLICATION NUMBER: PCT/US99/19330
; EARLIER FILING DATE: 1999-08-24
; EARLIER APPLICATION NUMBER: 60/097,917
; EARLIER FILING DATE: 1998-08-25
; EARLIER APPLICATION NUMBER: 60/098,634
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 170
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 61
; LENGTH: 1292
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (71)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (697)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1280)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1287)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1291)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-739-254-61

Query Match 81.9%; Score 1079; DB 10; Length 1292;
Best Local Similarity 97.3%; Pred. No. 2.1e-295;
Matches 1105; Conservative 4; Mismatches 20; Indels 7; Gaps 1;

```
Qy       182 AAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATCCAAGCCAACA 241
             ||  | |||||| |   |      |   | ||||||||||||||||||||||||||||||
Db        76 AAAGAGGAAACATAGAGGTGCCAAAGGAACAAAGACATAATGATGTCATCCAAGCCAACA 135

Qy       242 AGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAGCTTTATGGCT 301
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       136 AGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAGCTTTATGGCT 195

Qy       302 CCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCAGGGTGCTCAG 361
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       196 CCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCAGGGTGCTCAG 255

Qy       362 CGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCAACCGGGTCAA 421
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       256 CGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCAACCGGGTCAA 315
```

```
Qy        422 GGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTTTAAAGAAGAA 481
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 GGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTTTAAAGAAGAA 375

Qy        482 GCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGGTTTTGGAATT 541
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 GCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGGTTTTGGAATT 435

Qy        542 GTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCTACTGCTGTT 601
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db        436 GTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCTACTGCTGKT 495

Qy        602 ATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCTCTCTCTGTG 661
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 ATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCTCTCTCTGTG 555

Qy        662 TCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATGAACATTGTT 721
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        556 TCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATGAACATTGGT 615

Qy        722 AGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGCATCAATGGG 781
              || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 AGGTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGCATCAATGGG 675

Qy        782 GTAGCTGGCCAAGACTACTGGGCCGTGCTTTCTGGAAAAGGCATTTCAGCCACGCTGATG 841
              |||:|||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        676 GTARCTGGCCAAGACTACTGGNCCGTGCTTTCTGGAAAAGGCATTTCAGCCACGCTGATG 735

Qy        842 ATCTTCTCCCTCTTGGAGTTCTTCGTAGCTTGTGCCACAGCCCATTTTGCCAACCAAGCA 901
              |||||||||:||||||||||||||||||||||||||||||||||||||||||||||||||
Db        736 ATCTTCTCCYTCTTGGAGTTCTTCGTAGCTTGTGCCACAGCCCATTTTGCCAACCAAGCA 795

Qy        902 AACACCACAACCAATATGTCTGTCCTGGTTATTCCAAATATGTATGAAAGCAACCCTGTG 961
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db        796 AACACCACAACCAATATGTCTGTCCTGGTTATTCCAAATATGTATGAAAGCAACCCTGKG 855

Qy        962 ACACCAGCGTCTTCTTCAGCTCCTCCCAGATGCAACAACTACTCAGCTAATGCCCCTAAA 1021
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        856 ACACCAGCGTCTTCTTCAGCTCCTCCCAGATGCAACAACTACTCAGCTAATGCCCC---- 911

Qy       1022 TAGTAAAAGAAAAAGGGGTATCAGTCTAATCTCATGGAGAAAAACTACTTGCAAAAACTT 1081
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        912 ---TAAAAGAAAAAGGGGTATCAGTCTAATCTCATGGAGAAAAACTACTTGCAAAAACTT 968

Qy       1082 CTTAAGAAGATGTCTTTTATTGTCTACAATGATTTCTAGTCTTTAAAAACTGTGTTTGAG 1141
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         969 CTTAAGAAGATGTCTTTTATTGTCTACAATGATTTCTAGTCTTTAAAAACTGTGTTTGAG 1028

Qy       1142 ATTTGTTTTTAGGTTGGTCGCTAATGATGGCTGTATCTCCCTTCACTGTCTCTTCCTACA 1201
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1029 ATTTGTTTTTAGGTTGGTCGCTAATGATGGCTGTATCTCCCTTCACTGTCTCTTCCTACA 1088

Qy       1202 TTACCACTACTACATGCTGGCAAAGGTGAAGGATCAGAGGACTGAAAAATGATTCTGCAA 1261
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1089 TTACCACTACTACATGCTGGCAAAGGTGAAGGATCAGAGGACTGAAAAATGATTCTGCAA 1148

Qy       1262 CTCTCTTAAAGTTAGAAATGTTTCTGTTCATATTACTTTTTCCTTAATAAAATGTC 1317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1149 CTCTCTTAAAGTTAGAAATGTTTCTGTTCATATTACTTTTTCCTTAATAAAATGTC 1204
```

```
RESULT 8
US-09-904-615-61
; Sequence 61, Application US/09904615
; Patent No. US20020026040A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: 49 Human Secreted Proteins
;  FILE REFERENCE: PZ032P1
;  CURRENT APPLICATION NUMBER: US/09/904,615
;  CURRENT FILING DATE:  2001-07-16
;  PRIOR APPLICATION NUMBER: 09/511,554
;  PRIOR FILING DATE: 2000-02-23
;  PRIOR APPLICATION NUMBER: 60/097,917
;  PRIOR FILING DATE: 1998-08-25
;  PRIOR APPLICATION NUMBER: 60/098,634
;  PRIOR FILING DATE: 1998-08-31
;  NUMBER OF SEQ ID NOS: 170
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 61
;   LENGTH: 1292
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (71)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (697)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (1280)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (1287)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (1291)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-904-615-61

  Query Match             81.9%;  Score 1079;  DB 10;  Length 1292;
  Best Local Similarity   97.3%;  Pred. No. 2.1e-295;
  Matches 1105;  Conservative    4;  Mismatches   20;  Indels    7;  Gaps    1;

Qy        182 AAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATCCAAGCCAACA 241
              ||  | ||||||  |   |       |   | ||||||||||||||||||||||||||||
Db         76 AAAGAGGAAACATAGAGGTGCCAAAGGAACAAAGACATAATGATGTCATCCAAGCCAACA 135

Qy        242 AGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAGCTTTATGGCT 301
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 AGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAGCTTTATGGCT 195

Qy        302 CCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCAGGGTGCTCAG 361
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 CCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCAGGGTGCTCAG 255

Qy        362 CGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCAACCGGGTCAA 421
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 CGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCAACCGGGTCAA 315

Qy        422 GGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTTTAAAGAAGAA 481
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 GGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTTTAAAGAAGAA 375

Qy        482 GCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGGTTTTGGAATT 541
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 GCAAAGGCACTAGGGGTGATCCAGATCATGGTTGGATTGATGCACATTGGTTTTGGAATT 435

Qy        542 GTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCTACTGCTGTT 601
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db        436 GTTTTGTGTTTAATATCCTTCTCTTTTAGAGAAGTATTAGGTTTTGCCTCTACTGCTGKT 495

Qy        602 ATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCTCTCTCTGTG 661
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 ATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGCTCTCTCTCTGTG 555

Qy        662 TCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATGAACATTGTT 721
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        556 TCAGCATCCAAGGAGCTTTCCCGTTGTCTGGTGAAAGGCAGCCTGGGAATGAACATTGGT 615

Qy        722 AGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGCATCAATGGG 781
              || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 AGGTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGATATGTGCATCAATGGG 675

Qy        782 GTAGCTGGCCAAGACTACTGGGCCGTGCTTTCTGGAAAAGGCATTTCAGCCACGCTGATG 841
```

```
              |||:|||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        676 GTARCTGGCCAAGACTACTGGNCCGTGCTTTCTGGAAAAGGCATTTCAGCCACGCTGATG 735

Qy        842 ATCTTCTCCCTCTTGGAGTTCTTCGTAGCTTGTGCCACAGCCCATTTTGCCAACCAAGCA 901
              |||||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        736 ATCTTCTCCYTCTTGGAGTTCTTCGTAGCTTGTGCCACAGCCCATTTTGCCAACCAAGCA 795

Qy        902 AACACCACAACCAATATGTCTGTCCTGGTTATTCCAAATATGTATGAAAGCAACCCTGTG 961
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db        796 AACACCACAACCAATATGTCTGTCCTGGTTATTCCAAATATGTATGAAAGCAACCCTGKG 855

Qy        962 ACACCAGCGTCTTCTTCAGCTCCTCCCAGATGCAACAACTACTCAGCTAATGCCCCTAAA 1021
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        856 ACACCAGCGTCTTCTTCAGCTCCTCCCAGATGCAACAACTACTCAGCTAATGCCCC---- 911

Qy       1022 TAGTAAAAGAAAAAGGGGTATCAGTCTAATCTCATGGAGAAAAACTACTTGCAAAAACTT 1081
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        912 ---TAAAAGAAAAAGGGGTATCAGTCTAATCTCATGGAGAAAAACTACTTGCAAAAACTT 968

Qy       1082 CTTAAGAAGATGTCTTTTATTGTCTACAATGATTTCTAGTCTTTAAAAACTGTGTTTGAG 1141
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        969 CTTAAGAAGATGTCTTTTATTGTCTACAATGATTTCTAGTCTTTAAAAACTGTGTTTGAG 1028

Qy       1142 ATTTGTTTTTAGGTTGGTCGCTAATGATGGCTGTATCTCCCTTCACTGTCTCTTCCTACA 1201
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1029 ATTTGTTTTTAGGTTGGTCGCTAATGATGGCTGTATCTCCCTTCACTGTCTCTTCCTACA 1088

Qy       1202 TTACCACTACTACATGCTGGCAAAGGTGAAGGATCAGAGGACTGAAAAATGATTCTGCAA 1261
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1089 TTACCACTACTACATGCTGGCAAAGGTGAAGGATCAGAGGACTGAAAAATGATTCTGCAA 1148

Qy       1262 CTCTCTTAAAGTTAGAAATGTTTCTGTTCATATTACTTTTTCCTTAATAAAATGTC 1317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1149 CTCTCTTAAAGTTAGAAATGTTTCTGTTCATATTACTTTTTCCTTAATAAAATGTC 1204
```

```
RESULT 9
US-09-925-299-727
; Sequence 727, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA102
;  CURRENT APPLICATION NUMBER: US/09/925,299
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05883
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1556
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 727
;   LENGTH: 441
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (321)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (394)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (405)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (422)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (433)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (438)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-925-299-727

  Query Match             29.6%;  Score 389.6;  DB 9;  Length 441;
  Best Local Similarity   96.6%;  Pred. No. 3.9e-100;
  Matches  427;  Conservative    0;  Mismatches   10;  Indels    5;  Gaps    3;

Qy        113 GAAGACAAGTTCTTGACTCTATGTTGAGGCCAGTTGAAAAATGAGGGAGAATAAAACCAT 172
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 GAAGACAAGTTCTTGACTCTATGTTGAGGCCAGTTGAAAAATGAGGGAGAATAAAACCAT 61

Qy        173 GAACGAAACAAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATCC 232
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 GAACGAAACAAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATCC 121

Qy        233 AAGCCAACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAGC 292
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 AAGCCAACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAGC 181

Qy        293 TTTATGGCTCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCAG 352
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 TTTATGGCTCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCAG 241

Qy        353 GGTGCTCAGCGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCAA 412
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 GGTGCTCAGCGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCAA 301

Qy        413 CCGGGTCAAGGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTTT 472
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        302 CCGGGTCAAGGAAATATACNAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTTT 361

Qy        473 AAA--GAAGAAGCAAAGGCACTA-GGGGTGATCCAGATCATGGTTGGATTGATGCACATT 529
              |||   |||||||||||||||| |||||||| ||||||||| |||||||||  |  |
Db        362 AAAAGAAAGAAGCAAAGGCACTAGGGGGTGATNCAGATCATGGNTGGATTGATG--CCAT 419

Qy        530 GGTTTTGGAATTGTTTTGTGTT 551
               | |||||||||| |||| |||
Db        420 TGNTTTGGAATTGNTTTGNGTT 441
```

```
RESULT 10
US-09-925-299-727
; Sequence 727, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA102
;  CURRENT APPLICATION NUMBER: US/09/925,299
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05883
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1556
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 727
;   LENGTH: 441
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (321)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (394)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (405)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (422)
```

```
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (433)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (438)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-925-299-727

  Query Match            29.6%;  Score 389.6;  DB 10;  Length 441;
  Best Local Similarity  96.6%;  Pred. No. 3.9e-100;
  Matches  427;  Conservative    0;  Mismatches   10;  Indels    5;  Gaps    3;

Qy        113 GAAGACAAGTTCTTGACTCTATGTTGAGGCCAGTTGAAAAATGAGGGAGAATAAAACCAT 172
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 GAAGACAAGTTCTTGACTCTATGTTGAGGCCAGTTGAAAAATGAGGGAGAATAAAACCAT 61

Qy        173 GAACGAAACAAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATCC 232
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 GAACGAAACAAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATCC 121

Qy        233 AAGCCAACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAGC 292
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 AAGCCAACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAGC 181

Qy        293 TTTATGGCTCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCAG 352
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 TTTATGGCTCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCAG 241

Qy        353 GGTGCTCAGCGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCAA 412
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 GGTGCTCAGCGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCAA 301

Qy        413 CCGGGTCAAGGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTTT 472
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        302 CCGGGTCAAGGAAATATACNAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTTT 361

Qy        473 AAA--GAAGAAGCAAAGGCACTA-GGGGTGATCCAGATCATGGTTGGATTGATGCACATT 529
              |||   |||||||||||||||| |||||||| |||||||||| |||||||||| |  |
Db        362 AAAAGAAAGAAGCAAAGGCACTAGGGGGTGATNCAGATCATGGNTGGATTGATG--CCAT 419

Qy        530 GGTTTTGGAATTGTTTTGTGTT 551
               | |||||||||| |||| |||
Db        420 TGNTTTGGAATTGNTTTGNGTT 441


RESULT 11
US-09-803-719-494
; Sequence 494, Application US/09803719
; Publication No. US20030044783A1
; GENERAL INFORMATION:
;  APPLICANT: Williams, Lewis T.
;  APPLICANT:  Escobedo, Jaime
;  APPLICANT:  Innis, Michael A.
;  APPLICANT:  Garcia, Pablo Dominiguez
;  APPLICANT:  Sudduth-Klinger, Julie
;  APPLICANT:  Reinhard, Christoph
;  APPLICANT:  Giese, Klaus
;  APPLICANT:  Randazzo, Filippo
;  APPLICANT:  Kennedy, Giulia C.
;  APPLICANT:  Pot, David
;  APPLICANT:  Kassam, Altaf
;  APPLICANT:  Lamson, George
;  APPLICANT:  Drmanac, Radoje
;  APPLICANT:  Crkvenjakov, Radomir
;  APPLICANT:  Dickson, Mark
;  APPLICANT:  Drmanac, Snezana
;  APPLICANT:  Labat, Ivan
;  APPLICANT:  Leshkowitz, Dena
;  APPLICANT:  Kita, David
;  APPLICANT:  Garcia, Veronica
;  APPLICANT:  Jones, Lee William
;  APPLICANT:  Stache-Crain, Birgit
;  TITLE OF INVENTION: Human Genes and Gene Products
;  FILE REFERENCE: 1624.002
;  CURRENT APPLICATION NUMBER: US/09/803,719
;  CURRENT FILING DATE:  2001-03-09
;  PRIOR APPLICATION NUMBER: 60/188,609
;  PRIOR FILING DATE: 2000-03-09
;  NUMBER OF SEQ ID NOS: 2396
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 494
;   LENGTH: 351
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-803-719-494

  Query Match            23.1%;  Score 303.6;  DB 9;  Length 351;
  Best Local Similarity  93.0%;  Pred. No. 8.5e-76;
  Matches  318;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy        172 TGAACGAAACAAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATC 231
              || |  | | ||  | ||||||  |   |       |   | ||||||||||||||||||
Db         10 TGGAGCAGAGAAAGAGGAAACATAGAGGTGCCAAAGGAACAAAGACATAATGATGTCATC 69

Qy        232 CAAGCCAACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAG 291
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         70 CAAGCCAACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAG 129

Qy        292 CTTTATGGCTCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCA 351
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        130 CTTTATGGCTCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCA 189

Qy        352 GGGTGCTCAGCGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCA 411
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        190 GGGTGCTCAGCGTGCTCAGCCCTATGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCA 249

Qy        412 ACCGGGTCAAGGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTT 471
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        250 ACCGGGTCAAGGAAATATATAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTT 309

Qy        472 TAAAGAAGAAGCAAAGGCACTAGGGGTGATCCAGATCATGGT 513
              |||||||||||||||||||||||||||||||||||||||||||
Db        310 TAAAGAAGAAGCAAAGGCACTAGGGGTGATCCAGATCATGGT 351


RESULT 12
US-09-803-719-638
; Sequence 638, Application US/09803719
; Publication No. US20030044783A1
; GENERAL INFORMATION:
;  APPLICANT: Williams, Lewis T.
;  APPLICANT:  Escobedo, Jaime
;  APPLICANT:  Innis, Michael A.
;  APPLICANT:  Garcia, Pablo Dominiguez
;  APPLICANT:  Sudduth-Klinger, Julie
;  APPLICANT:  Reinhard, Christoph
;  APPLICANT:  Giese, Klaus
;  APPLICANT:  Randazzo, Filippo
;  APPLICANT:  Kennedy, Giulia C.
;  APPLICANT:  Pot, David
;  APPLICANT:  Kassam, Altaf
;  APPLICANT:  Lamson, George
;  APPLICANT:  Drmanac, Radoje
;  APPLICANT:  Crkvenjakov, Radomir
;  APPLICANT:  Dickson, Mark
;  APPLICANT:  Drmanac, Snezana
;  APPLICANT:  Labat, Ivan
;  APPLICANT:  Leshkowitz, Dena
;  APPLICANT:  Kita, David
;  APPLICANT:  Garcia, Veronica
;  APPLICANT:  Jones, Lee William
;  APPLICANT:  Stache-Crain, Birgit
;  TITLE OF INVENTION: Human Genes and Gene Products
;  FILE REFERENCE: 1624.002
;  CURRENT APPLICATION NUMBER: US/09/803,719
```

```
;  CURRENT FILING DATE:  2001-03-09
;  PRIOR APPLICATION NUMBER: 60/188,609
;  PRIOR FILING DATE: 2000-03-09
;  NUMBER OF SEQ ID NOS: 2396
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 638
;   LENGTH: 342
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(342)
;   OTHER INFORMATION: n = A,T,C or G
US-09-803-719-638

  Query Match            22.2%;  Score 292.6;  DB 9;  Length 342;
  Best Local Similarity   92.5%;  Pred. No. 1.1e-72;
  Matches  307;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy        172 TGAACGAAACAAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATC 231
              || | | | ||  | |||||| |   |      |   | ||||||||||||||||||||
Db         10 TGGAGCAGAGAAAGAGGAAACATAGAGGTGCCAAAGGAACAAAGACATAATGATGTCATC 69

Qy        232 CAAGCCAACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAG 291
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         70 CAAGCCAACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAG 129

Qy        292 CTTTATGGCTCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCA 351
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        130 CTTTATGGCTCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCA 189

Qy        352 GGGTGCTCAGCGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCA 411
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db        190 GGGTGCTCAGCGTGCTCAGCCCTATGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCA 249

Qy        412 ACCGGGTCAAGGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTT 471
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        250 ACCGGGTCAAGGAAATATATAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTT 309

Qy        472 TAAAGAAGAAGCAAAGGCACTAGGGGTGATCC 503
              |||||||||||||  ||||||||||||||||||
Db        310 TAAAGAAGAAGCANAGGCACTAGGGGTGATCC 341


RESULT 13
US-09-803-719-481
; Sequence 481, Application US/09803719
; Publication No. US20030044783A1
; GENERAL INFORMATION:
;  APPLICANT: Williams, Lewis T.
;  APPLICANT:  Escobedo, Jaime
;  APPLICANT:  Innis, Michael A.
;  APPLICANT:  Garcia, Pablo Dominiguez
;  APPLICANT:  Sudduth-Klinger, Julie
;  APPLICANT:  Reinhard, Christoph
;  APPLICANT:  Giese, Klaus
;  APPLICANT:  Randazzo, Filippo
;  APPLICANT:  Kennedy, Giulia C.
;  APPLICANT:  Pot, David
;  APPLICANT:  Kassam, Altaf
;  APPLICANT:  Lamson, George
;  APPLICANT:  Drmanac, Radoje
;  APPLICANT:  Crkvenjakov, Radomir
;  APPLICANT:  Dickson, Mark
;  APPLICANT:  Drmanac, Snezana
;  APPLICANT:  Labat, Ivan
;  APPLICANT:  Leshkowitz, Dena
;  APPLICANT:  Kita, David
;  APPLICANT:  Garcia, Veronica
;  APPLICANT:  Jones, Lee William
;  APPLICANT:  Stache-Crain, Birgit
;  TITLE OF INVENTION: Human Genes and Gene Products
;  FILE REFERENCE: 1624.002
;  CURRENT APPLICATION NUMBER: US/09/803,719
;  CURRENT FILING DATE:  2001-03-09
;  PRIOR APPLICATION NUMBER: 60/188,609
;  PRIOR FILING DATE: 2000-03-09
;  NUMBER OF SEQ ID NOS: 2396
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 481
;   LENGTH: 383
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(383)
;   OTHER INFORMATION: n = A,T,C or G
US-09-803-719-481

  Query Match            21.9%;  Score 288.6;  DB 9;  Length 383;
  Best Local Similarity   90.1%;  Pred. No. 1.6e-71;
  Matches  309;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy        172 TGAACGAAACAAGAAAGAAACAAAACAGAAGAGGAATGAAAAAGACATAATGATGTCATC 231
              || |  | | ||  ||||| || |   |      |   |  ||||||||||||||| |
Db         41 TGGAGCACAGAAAGAAGAATCATAGAGGTGCCAAAGGAACTTAGACATAATGATGTCGTT 100

Qy        232 CAAGCCAACAAGCCATGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAG 291
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        101 CAAGCCAACAAGCCAAGCTGAAGTAAATGAAACCATACCCAACCCTTACCCACCAAGCAG 160

Qy        292 CTTTATGGCTCCTGGATTTCAACAGCCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCA 351
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        161 TTTTATGGCTCCTGGATTTCAACAGGCTCTGGGTTCAATCAACTTAGAAAACCAAGCTCA 220

Qy        352 GGGTGCTCAGCGTGCTCAGCCCTACGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCA 411
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        221 TGGTGCTCAGCGTGCTCAGCCCTATGGCATCACATCTCCGGGAATCTTTGCTAGCAGTCA 280

Qy        412 ACCGGGTCAAGGAAATATACAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTT 471
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        281 ACCGGGTCAAGGAAATATATAAATGATAAATCCAAGTGTGGGAACAGCAGTAATGAACTT 340

Qy        472 TAAAGAAGAAGCAAAGGCACTAGGGGTGATCCAGATCATGGTT 514
              |||||||||||||||||||||| ||||||||| |||||||||
Db        341 TAAAGAAGAAGCAAAGGCACTATGGGTGATCCACATCATGGTT 383


RESULT 14
US-09-563-817-780
; Sequence 780, Application US/09563817
; Patent No. US20020095031A1
; GENERAL INFORMATION:
;  APPLICANT: Nehls, Michael C.
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. US20020095031Alel Human Polynucleotides and the
;  TITLE OF INVENTION:  Polypeptides Encoded Thereby
;  FILE REFERENCE: LEX-0021-USA
;  CURRENT APPLICATION NUMBER: US/09/563,817
;  CURRENT FILING DATE:  2001-06-11
;  PRIOR APPLICATION NUMBER: US 60/132,343
;  PRIOR FILING DATE: 1999-05-04
;  NUMBER OF SEQ ID NOS: 1008
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 780
;   LENGTH: 125
;   TYPE: DNA
;   ORGANISM: homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(125)
;   OTHER INFORMATION: n = A,T,C or G
US-09-563-817-780

  Query Match             7.0%;  Score 92.6;  DB 10;  Length 125;
```

```
  Best Local Similarity   87.2%;  Pred. No. 3.1e-16;
  Matches  109;  Conservative    0;  Mismatches   15;  Indels    1;  Gaps    1;

Qy        508 CATGGTTGGATTGATGCACATTGGT-TTTGGAATTGTTTTGTGTTTAATATCCTTCTCTT 566
              | ||||||||||||||||| |||||  | || |  |   || ||  ||||||||||||
Db          1 CGTGGTTGGATTGATGCACNTTGGTGCTCGGNAACGCACTGNGTNNAATATCCTTCTCTT 60

Qy        567 TTAGAGAAGTATTAGGTTTTGCCTCTACTGCTGTTATTGGTGGATACCCATTCTGGGGTG 626
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||| |
Db         61 TTAGAGAAGTATTANGTTTTGCCTCTACTGCTGTTATTGGTGGATACCCATTCTGGGGGG 120

Qy        627 GCCTT 631
              |||||
Db        121 GCCTT 125


RESULT 15
US-10-017-754-2003
; Sequence 2003, Application US/10017754
; Publication No. US20030054363A1
; GENERAL INFORMATION:
;  APPLICANT: Henderson, Robert A.
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Watanabe, Yoshihiro
;  APPLICANT:  Johnson, Jeffrey C.
;  APPLICANT:  Retter, Marc W.
;  APPLICANT:  Marnerakis, Margarita
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Vedvick, Thomas S.
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  McNabb, Andria
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.478C18
;  CURRENT APPLICATION NUMBER: US/10/017,754
;  CURRENT FILING DATE:  2001-10-29
;  NUMBER OF SEQ ID NOS: 2004
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2003
;   LENGTH: 753
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-017-754-2003

  Query Match             6.2%;  Score 81.2;  DB 9;  Length 753;
  Best Local Similarity   58.3%;  Pred. No. 1.8e-12;
  Matches  190;  Conservative    0;  Mismatches  118;  Indels   18;  Gaps    2;

Qy        590 TCTACTGCTGTTATTGGTGGATACCCATTCTGGGGTGGCCTTTCTTTTATTATCTCTGGC 649
              |||| | |  |    || || | || |||||||| ||| | |  ||||| || || ||
Db        298 TCTATTTCATTCTACGGAGGCTTTCCCTTCTGGGGAGGCTTGTGGTTTATCATTTCAGGA 357

Qy        650 TCTCTCTCTGTGTCAGCATCCAAGGAGCT---TTCCCGTTGTCTGGTGAAAGGCAGCCTG 706
              |||||||| ||| |||||   ||  |||   |||  ||| ||| ||  |||||  ||
Db        358 TCTCTCTCCGTGGCAGCAGAAAATCAGCCATATTCTTATTGCCTGCTGTCTGGCAGTTTG 417

Qy        707 GGAATGAACATTGTTAGTTCTATCTTGGCCTTCATTGGAGTGATTCTGCTGCTGGTGGAT 766
              ||  ||||||| || ||| | ||||   |     ||||||| || ||  |  |    |||
Db        418 GGCTTGAACATCGTCAGTGCAATCTGCTCTGCAGTTGGAGTCATACTCTTCATCACAGAT 477

Qy        767 ATGTGCAT---------------CAATGGGGTAGCTGGCCAAGACTACTGGGCCGTGCTT 811
               |  | ||                | |   |   |   |  ||  |||||  |||
Db        478 CTAAGTATTCCCCACCCATATGCCTACCCCGACTATTATCCTTACGCCTGGGGTGTGAAC 537

Qy        812 TCTGGAAAAGGCATTTCAGCCACGCTGATGATCTTCTCCCTCTTGGAGTTCTTCGTAGCT 871
              ||||||  |  ||||| | |  |||| || |||||| |||| |||||||   | | ||
Db        538 CCTGGAATGGCGATTTCTGGCGTGCTGCTGGTCTTCTGCCTCCTGGAGTTTGGCATCGCA 597

Qy        872 TGTGCCACAGCCCATTTTGCCAACCA 897
              || ||  |  |||| |||| |  |||
Db        598 TGCGCATCTTCCCACTTTGGCTGCCA 623
```

```
Search completed: July 16, 2003, 18:04:17
Job time : 176.297 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 16, 2003, 03:49:46 ; Search time 1060.5 Seconds
(without alignments)
15998.962 Million cell updates/sec

Title: US-09-981-353-114
Perfect score: 583
Sequence: 1 ccctgtaggtgctgtccttg.....tatgtgaatgataatagtca 583

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		304.4	52.2	465	9	BC029475	BC029475 Homo sapi
2		277.8	47.7	454	9	BC032338	BC032338 Homo sapi
3		277.4	47.6	415	6	AX330310	AX330310 Sequence
4		277.4	47.6	415	6	AX336406	AX336406 Sequence
5		277.4	47.6	415	6	AX411150	AX411150 Sequence
6		277.4	47.6	415	9	HSMT1L	X76717 H.sapiens M
7		276.2	47.4	404	9	BC018190	BC018190 Homo sapi
8		260.2	44.6	120825	9	HSJ733M16	AL109627 Human DNA
9		259.8	44.6	368	9	AY077636	AY077636 Homo sapi
10		248	42.5	72930	9	HS1174N9	AL031602 Human DNA
11		246.4	42.3	422	9	BC008408	BC008408 Homo sapi
12	c	245.2	42.1	409	6	AX330235	AX330235 Sequence
13	c	245.2	42.1	409	6	AX331811	AX331811 Sequence
14	c	245.2	42.1	409	6	AX336059	AX336059 Sequence
15	c	245.2	42.1	409	6	AX336280	AX336280 Sequence
16	c	245.2	42.1	409	6	AX336680	AX336680 Sequence
17	c	245.2	42.1	409	6	AX410056	AX410056 Sequence
18		237	40.7	844	9	HUMMETIPG	M13073 Human metal
19		234.4	40.2	729	9	BC009699	BC009699 Homo sapi
20		233.2	40.0	288	11	G06180	G06180 human STS W
21		227.2	39.0	468	9	BC029453	BC029453 Homo sapi
22		226.8	38.9	1071	9	AF078844	AF078844 Homo sapi
23		226.6	38.9	164529	2	AC079975	AC079975 Homo sapi
24	c	226.6	38.9	164933	2	AC074231	AC074231 Homo sapi
25	c	226.6	38.9	172753	9	AL359259	AL359259 Human DNA
26		226.6	38.9	176671	2	AC080149	AC080149 Homo sapi
27		221.8	38.0	434	9	BC020757	BC020757 Homo sapi
28		221.8	38.0	453	6	AX399847	AX399847 Sequence
29		219.8	37.7	367	6	AX336464	AX336464 Sequence
30		219.8	37.7	367	6	AX411116	AX411116 Sequence
31		219.8	37.7	367	9	HSFMET	X64177 H.sapiens m
32		219.4	37.6	431	9	BC028280	BC028280 Homo sapi
33	c	219.4	37.6	160716	2	AC053475	AC053475 Homo sapi
34		219.4	37.6	176668	2	AC027096	AC027096 Homo sapi
35		219.4	37.6	189643	9	AL161729	AL161729 Human DNA
36		216	37.0	646	6	AX067316	AX067316 Sequence
37	c	216	37.0	738	9	AB064124	AB064124 Homo sapi
38	c	216	37.0	761	9	AB064049	AB064049 Homo sapi
39	c	216	37.0	781	9	AB064131	AB064131 Homo sapi
40	c	216	37.0	786	9	AB064060	AB064060 Homo sapi
41	c	216	37.0	791	9	AB064135	AB064135 Homo sapi
42	c	216	37.0	794	9	AB064122	AB064122 Homo sapi
43	c	216	37.0	794	9	AB064129	AB064129 Homo sapi
44	c	216	37.0	796	9	AB064120	AB064120 Homo sapi
45	c	216	37.0	796	9	AB064128	AB064128 Homo sapi

ALIGNMENTS

RESULT 1
BC029475
LOCUS BC029475 465 bp mRNA linear PRI 16-MAY-2002
DEFINITION Homo sapiens, Similar to RNA helicase-related protein, clone MGC:32848 IMAGE:4723818, mRNA, complete cds.
ACCESSION BC029475
VERSION BC029475.1 GI:20809974
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 465)
AUTHORS Strausberg,R.
TITLE Direct Submission

```
  JOURNAL   Submitted (01-MAY-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:       http://www-shgc.stanford.edu
            Contact:  (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 41 Row: m Column: 3
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis.
FEATURES             Location/Qualifiers
     source          1. .465
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="MGC:32848 IMAGE:4723818"
                     /tissue_type="Liver"
                     /clone_lib="NIH_MGC_76"
                     /lab_host="DH10B"
                     /note="Vector: pDNR-LIB"
     CDS             71. .256
                     /codon_start=1
                     /product="Similar to RNA helicase-related protein"
                     /protein_id="AAH29475.1"
                     /db_xref="GI:20809975"
                     /translation="MDPNCSCATGGSCTCTGSCKCKECKCNSCKKSCCSCCPMSCAKC
                     AQGCICKGASEKCSCCA"
BASE COUNT      169 a    109 c     88 g     99 t
ORIGIN

  Query Match           52.2%;  Score 304.4;  DB 9;  Length 465;
  Best Local Similarity   91.2%;  Pred. No. 2.5e-74;
  Matches  323;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy        226 GCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCCCAACT 285
              || | |.  | ||||| | |||| ||||  | || ||  | ||||||||||||||||||
Db         24 GCCTCTCAACTTCTTGCTTGGGATCTCCAACCTCACCGCGGCTCGAAATGGACCCCAACT 83

Qy        286 GCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGAGTGCA 345
              |||||||| |  ||| |||||||||  ||||  | ||||||||||||||||||||||
Db         84 GCTCCTGCGCCACTGGTGGCTCCTGCACCTGCACTGGCTCCTGCAAATGCAAAGAGTGCA 143

Qy        346 AATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAAGTGTG 405
              |||||| ||||||||||||||||||||||||||||||  || ||||||||||||||||
Db        144 AATGCAACTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCCATGAGCTGTGCCAAGTGTG 203

Qy        406 CCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATGTCCGG 465
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        204 CCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATGTCCGG 263

Qy        466 ACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGAATTTTTTTTCCAT 525
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        264 ACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGAATTTTTTTTCCAT 323

Qy        526 ACAACCCTGCCCCCATCTACTGTATTTTTTTTAATGAAATATGTGAATGATAATA 579
              |||||||| ||||| ||||||||||||||||||||||||||||||||||||||
Db        324 ACAACCCTGACCCATTTACTGTATTTTTTTTAATGAAATATGTGAATGATAATA 377


RESULT 2
BC032338
LOCUS       BC032338                454 bp    mRNA    linear   PRI 27-JUN-2002
DEFINITION  Homo sapiens, metallothionein 1L, clone MGC:40349 IMAGE:5190391,
            mRNA, complete cds.
ACCESSION   BC032338
VERSION     BC032338.1  GI:21619585
KEYWORDS    MGC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 454)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUN-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
            Web site: http://www.nisc.nih.gov/
            Contact: nisc_mgc@nhgri.nih.gov
            Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
            Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
            Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D.,McCloskey,J.C.,
            McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
            Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
            Young,A., Zhang,L.-H. and Green,E.D.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 64 Row: g Column: 18
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 4505270.
FEATURES             Location/Qualifiers
     source          1. .454
                     /organism="Homo sapiens"
                     /db_xref="LocusID:4500"
                     /db_xref="taxon:9606"
                     /clone="MGC:40349 IMAGE:5190391"
                     /tissue_type="Colon, Kidney, Stomach, adult, whole pooled"
                     /clone_lib="NIH_MGC_116"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS             59. .244
                     /codon_start=1
                     /product="metallothionein 1L"
                     /protein_id="AAH32338.1"
                     /db_xref="GI:21619586"
                     /translation="MDPNCSCSPVGSCACAGSCKCKECKCTSCKKSCCSCCPVGCAKC
                     AQGCICKGTSDKCSCCA"
BASE COUNT      143 a    105 c     88 g    118 t
ORIGIN

  Query Match           47.7%;  Score 277.8;  DB 9;  Length 454;
  Best Local Similarity   88.6%;  Pred. No. 7e-67;
  Matches  327;  Conservative    0;  Mismatches   32;  Indels   10;  Gaps    2;

Qy        221 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          7 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 66

Qy        281 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         67 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 126
```

```
Qy        341 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAA 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        127 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAA 186

Qy        401 GTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATG 460
              |||||||||||||||||||||||||||| | ||||| ||||||||||||||||||||||
Db        187 GTGTGCCCAGGGCTGCATCTGCAAAGGGACGTCAGACAAGTGCAGCTGCTGTGCCTGATG 246

Qy        461 TCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGA-------- 512
              | ||||||| ||||| |||| || | |||| |||| | |||| ||||
Db        247 CCAGGACAGCTGTGCTCTCAGATGTAAATAGAGCAACCTATATAAACCTGGATTTTTTTT 306

Qy        513 ATTTTTTTTCCATACAACCCTGCCCCATCTACT--GTATTTTTTTTAATGAAATATGTGA 570
              ||||||||   |||||||||| ||| | | ||   | |||||||  ||||||||||||
Db        307 TTTTTTTTTTGTACAACCCTGACCCGTTTGCTACATCTTTTTTTCTATGAAATATGTGA 366

Qy        571 ATGATAATA 579
              |||  ||||
Db        367 ATGGCAATA 375
```

```
RESULT 3
AX330310
LOCUS       AX330310                415 bp    DNA     linear   PAT 09-JAN-2002
DEFINITION  Sequence 819 from Patent WO0194629.
ACCESSION   AX330310
VERSION     AX330310.1  GI:18103288
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
            Horrigan,S., Soppet,D.R. and Weaver,Z.
  TITLE     Cancer gene determination and therapeutic screening using signature
            gene sets
  JOURNAL   Patent: WO 0194629-A 819 13-DEC-2001;
            Avalon Pharmaceuticals (US)
FEATURES             Location/Qualifiers
     source          1. .415
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      110 a    102 c     87 g    115 t      1 others
ORIGIN

  Query Match           47.6%;  Score 277.4;  DB 6;  Length 415;
  Best Local Similarity  88.3%;  Pred. No. 9e-67;
  Matches  326;  Conservative    1;  Mismatches   32;  Indels   10;  Gaps    2;

Qy        221 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 60

Qy        281 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 120

Qy        341 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAA 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:||
Db        121 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCMAA 180

Qy        401 GTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATG 460
              |||||||||||||||||||||||||||| | ||||| ||||||||||||||||||||||
Db        181 GTGTGCCCAGGGCTGCATCTGCAAAGGGACGTCAGACAAGTGCAGCTGCTGTGCCTGATG 240

Qy        461 TCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGA-------- 512
              | ||||||| ||||| |||| || | |||| |||| | |||| ||||
Db        241 CCAGGACAGCTGTGCTCTCAGATGTAAATAGAGCAACCTATATAAACCTGGATTTTTTTT 300

Qy        513 ATTTTTTTTCCATACAACCCTGCCCCATCTACT--GTATTTTTTTTAATGAAATATGTGA 570
              ||||||||   |||||||||| ||| | | ||   | |||||||  ||||||||||||
Db        301 TTTTTTTTTTGTACAACCCTGACCCGTTTGCTACATCTTTTTTTCTATGAAATATGTGA 360

Qy        571 ATGATAATA 579
              |||  ||||
Db        361 ATGGCAATA 369
```

```
RESULT 4
AX336406
LOCUS       AX336406                415 bp    DNA     linear   PAT 09-JAN-2002
DEFINITION  Sequence 6915 from Patent WO0194629.
ACCESSION   AX336406
VERSION     AX336406.1  GI:18127125
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
            Horrigan,S., Soppet,D.R. and Weaver,Z.
  TITLE     Cancer gene determination and therapeutic screening using signature
            gene sets
  JOURNAL   Patent: WO 0194629-A 6915 13-DEC-2001;
            Avalon Pharmaceuticals (US)
FEATURES             Location/Qualifiers
     source          1. .415
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      110 a    102 c     87 g    115 t      1 others
ORIGIN

  Query Match           47.6%;  Score 277.4;  DB 6;  Length 415;
  Best Local Similarity  88.3%;  Pred. No. 9e-67;
  Matches  326;  Conservative    1;  Mismatches   32;  Indels   10;  Gaps    2;

Qy        221 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 60

Qy        281 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 120

Qy        341 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAA 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:||
Db        121 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCMAA 180

Qy        401 GTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATG 460
              |||||||||||||||||||||||||||| | ||||| ||||||||||||||||||||||
Db        181 GTGTGCCCAGGGCTGCATCTGCAAAGGGACGTCAGACAAGTGCAGCTGCTGTGCCTGATG 240

Qy        461 TCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGA-------- 512
              | ||||||| ||||| |||| || | |||| |||| | |||| ||||
Db        241 CCAGGACAGCTGTGCTCTCAGATGTAAATAGAGCAACCTATATAAACCTGGATTTTTTTT 300

Qy        513 ATTTTTTTTCCATACAACCCTGCCCCATCTACT--GTATTTTTTTTAATGAAATATGTGA 570
              ||||||||   |||||||||| ||| | | ||   | |||||||  ||||||||||||
Db        301 TTTTTTTTTTGTACAACCCTGACCCGTTTGCTACATCTTTTTTTCTATGAAATATGTGA 360

Qy        571 ATGATAATA 579
              |||  ||||
Db        361 ATGGCAATA 369
```

```
RESULT 5
AX411150
LOCUS       AX411150                415 bp    DNA     linear   PAT 14-JUN-2002
DEFINITION  Sequence 3797 from Patent WO0229103.
ACCESSION   AX411150
VERSION     AX411150.1  GI:21443855
KEYWORDS    .
SOURCE      human.
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Alvares,C., Horne,D., Peres-da-Silva,S. and Vockley,J.G.
  TITLE     Gene expression profiles in liver cancer
  JOURNAL   Patent: WO 0229103-A 3797 11-APR-2002;
            GENE LOGIC INC (US)
FEATURES             Location/Qualifiers
     source          1. .415
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /note="EMBL/GenBank Accession No. X76717"
BASE COUNT      110 a    102 c     87 g    115 t      1 others
ORIGIN

  Query Match             47.6%;  Score 277.4;  DB 6;  Length 415;
  Best Local Similarity   88.3%;  Pred. No. 9e-67;
  Matches  326;  Conservative    1;  Mismatches   32;  Indels   10;  Gaps    2;

Qy        221 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 60

Qy        281 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 120

Qy        341 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAA 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:||
Db        121 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCMAA 180

Qy        401 GTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATG 460
              |||||||||||||||||||||||||| | ||||| ||||||||||||||||||||||||
Db        181 GTGTGCCCAGGGCTGCATCTGCAAAGGGACGTCAGACAAGTGCAGCTGCTGTGCCTGATG 240

Qy        461 TCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGA-------- 512
              | |||||||  |||||  |||| || | ||||  ||||  | |||| ||||
Db        241 CCAGGACAGCTGTGCTCTCAGATGTAAATAGAGCAACCTATATAAACCTGGATTTTTTTT 300

Qy        513 ATTTTTTTTCCATACAACCCTGCCCCATCTACT--GTATTTTTTTTAATGAAATATGTGA 570
              ||||||||   |||||||||| ||| | | ||   | |||||||  ||||||||||||
Db        301 TTTTTTTTTTTGTACAACCCTGACCCGTTTGCTACATCTTTTTTTCTATGAAATATGTGA 360

Qy        571 ATGATAATA 579
              |||  ||||
Db        361 ATGGCAATA 369


RESULT 6
HSMT1L
LOCUS       HSMT1L                   415 bp    mRNA    linear   PRI 10-JAN-1994
DEFINITION  H.sapiens MT-1l mRNA.
ACCESSION   X76717
VERSION     X76717.1  GI:435674
KEYWORDS    MT-1l protein.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 415)
  AUTHORS   Soumillion,A., Van Weyenbergh,J. and De Ley,M.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 415)
  AUTHORS   De Ley,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-1993) M. De Ley, Katholieke Universiteit Leuven,
            Lab voor Biochemie, Dekenstraat 6, 3000 Leuven, BELGIUM
FEATURES             Location/Qualifiers
     source          1. .415
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="pBlue-MT-11"
                     /tissue_type="liver"
                     /clone_lib="lambdaZAP-MT-1l"
                     /dev_stage="fetus"
     CDS             53. .238
                     /codon_start=1
                     /product="MT-1l protein"
                     /protein_id="CAA54136.1"
                     /db_xref="GI:435675"
                     /db_xref="SWISS-PROT:P80297"
                     /translation="MDPNCSCSPVGSCACAGSCKCKECKCTSCKKSCCSCCPVGCAKC
                     AQGCICKGTSDKCSCCA"
     polyA_signal    366. .371
BASE COUNT      110 a    102 c     87 g    115 t      1 others
ORIGIN

  Query Match             47.6%;  Score 277.4;  DB 9;  Length 415;
  Best Local Similarity   88.3%;  Pred. No. 9e-67;
  Matches  326;  Conservative    1;  Mismatches   32;  Indels   10;  Gaps    2;

Qy        221 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 60

Qy        281 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 120

Qy        341 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAA 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:||
Db        121 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCMAA 180

Qy        401 GTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATG 460
              |||||||||||||||||||||||||| | ||||| ||||||||||||||||||||||||
Db        181 GTGTGCCCAGGGCTGCATCTGCAAAGGGACGTCAGACAAGTGCAGCTGCTGTGCCTGATG 240

Qy        461 TCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGA-------- 512
              | |||||||  |||||  |||| || | ||||  ||||  | |||| ||||
Db        241 CCAGGACAGCTGTGCTCTCAGATGTAAATAGAGCAACCTATATAAACCTGGATTTTTTTT 300

Qy        513 ATTTTTTTTCCATACAACCCTGCCCCATCTACT--GTATTTTTTTTAATGAAATATGTGA 570
              ||||||||   |||||||||| ||| | | ||   | |||||||  ||||||||||||
Db        301 TTTTTTTTTTTGTACAACCCTGACCCGTTTGCTACATCTTTTTTTCTATGAAATATGTGA 360

Qy        571 ATGATAATA 579
              |||  ||||
Db        361 ATGGCAATA 369


RESULT 7
BC018190
LOCUS       BC018190                 404 bp    mRNA    linear   PRI 06-DEC-2001
DEFINITION  Homo sapiens, Similar to metallothionein 1L, clone MGC:9187
            IMAGE:3859643, mRNA, complete cds.
ACCESSION   BC018190
VERSION     BC018190.1  GI:17390422
KEYWORDS    MGC.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 404)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-DEC-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
```

```
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: villalon@bcm.tmc.edu.
            Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
            A.M., Holloway, M., Telford, B, Hodgson, A., Bouck, J., Yu, W.,
            Muzny,D.M., Gibbs,R.A.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 21 Row: d Column: 7
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 4505270.
FEATURES             Location/Qualifiers
     source          1. .404
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="MGC:9187 IMAGE:3859643"
                     /tissue_type="Ovary, adenocarcinoma"
                     /clone_lib="NIH_MGC_66"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS             57. .242
                     /codon_start=1
                     /product="Similar to metallothionein 1L"
                     /protein_id="AAH18190.1"
                     /db_xref="GI:17390423"
                     /translation="MDPNCSCSPVGSCACAGSCKCNECKCTSCKKSCCSCCPVGCAKC
                     AQGCICKGTSDKCSCCA"
BASE COUNT       94 a    105 c     88 g    117 t
ORIGIN

  Query Match           47.4%;  Score 276.2;  DB 9;  Length 404;
  Best Local Similarity  88.3%;  Pred. No. 1.9e-66;
  Matches  326;  Conservative    0;  Mismatches   33;  Indels   10;  Gaps    2;

Qy        221 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          5 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 64

Qy        281 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 340
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db         65 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAACGA 124

Qy        341 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAA 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        125 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAA 184

Qy        401 GTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATG 460
              |||||||||||||||||||||||||| | ||||| ||||||||||||||||||||||||
Db        185 GTGTGCCCAGGGCTGCATCTGCAAAGGGACGTCAGACAAGTGCAGCTGCTGTGCCTGATG 244

Qy        461 TCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGA-------- 512
               | |||||||  |||||  |||| || | ||||  ||||  | |||| ||||
Db        245 CCAGGACAGCTGTGCTCTCAGATGTAAATAGAGCAACCTATATAAACCTGGATTTTTTTT 304

Qy        513 ATTTTTTTTCCATACAACCCTGCCCCATCTACT--GTATTTTTTTTAATGAAATATGTGA 570
               ||||||||   ||||||||| ||| | | ||    | |||||||  ||||||||||||
Db        305 TTTTTTTTTTTGTACAACCCTGACCCGTTTGCTACATCTTTTTTTCTATGAAATATGTGA 364

Qy        571 ATGATAATA 579
              |||  ||||
Db        365 ATGGCAATA 373


RESULT 8
HSJ733M16
LOCUS       HSJ733M16              120825 bp    DNA     linear   PRI 06-DEC-2000
DEFINITION  Human DNA sequence from clone RP4-733M16 on chromosome
            1p36.11-36.23 Contains a pseudogene similar to metallothionein 1L,
            an EST, STSs, GSSs and CpG Islands, complete sequence.
ACCESSION   AL109627
VERSION     AL109627.18  GI:6742145
KEYWORDS    HTG; CpG Island; metallothionein.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 120825)
  AUTHORS   Hall,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jan 24, 2000 this sequence version replaced gi:6723665.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence is
            the entire insert of clone RP4-733M16 This sequence has been
            finished according to sequence map criteria as follows.  An attempt
            is made to resolve all sequencing problems, such as compressions
            and repeats, but not necessarily within known annotated repeat
            sequence elements.  Where the sequence is ambiguous, there is an
            annotation using the 'unsure' feature key. This sequence was
            generated from part of bacterial clone contigs of human chromosome
            1, constructed by the Sanger Centre Chromosome 1 Mapping Group.
            Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP4-733M16 is from the library RPCI-4 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2.
FEATURES             Location/Qualifiers
     source          1. .120825
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="p36.11-36.23"
                     /clone="RP4-733M16"
                     /clone_lib="RPCI-4"
     repeat_region   238. .444
                     /note="MER75 repeat: matches 1. .208 of consensus"
     repeat_region   2440. .2713
                     /note="AluSg1 repeat: matches 1. .275 of consensus"
     repeat_region   3522. .3597
                     /note="L1MC1 repeat: matches 6256. .6331 of consensus"
     repeat_region   3902. .4310
                     /note="L1MC1 repeat: matches 5848. .6256 of consensus"
     repeat_region   4613. .5146
                     /note="L1MC1 repeat: matches 5313. .5848 of consensus"
     repeat_region   5147. .5184
                     /note="19 copies 2 mer ta 94% conserved"
     repeat_region   5645. .5695
                     /note="L1MC1 repeat: matches 5276. .5330 of consensus"
     repeat_region   5977. .6150
                     /note="L1MC1 repeat: matches 5127. .5276 of consensus"
     repeat_region   7076. .7297
                     /note="L1MC1 repeat: matches 4890. .5132 of consensus"
     repeat_region   7313. .7633
                     /note="AluYa8 repeat: matches 1. .310 of consensus"
     repeat_region   8356. .8408
                     /note="L1MC1 repeat: matches 4850. .4900 of consensus"
     repeat_region   8538. .9097
                     /note="L1MC1 repeat: matches 4264. .4850 of consensus"
     misc_feature    complement(9794. .10350)
                     /note="match: GSS: Em:AQ535906"
     repeat_region   9824. .9994
```

```
                /note="AluSg/x repeat: matches 134. .305 of consensus"
repeat_region   13703. .13879
                /note="THE1C repeat: matches 2. .180 of consensus"
repeat_region   14176. .14372
                /note="THE1C repeat: matches 180. .371 of consensus"
repeat_region   14800. .14856
                /note="AluJ/FLAM repeat: matches 44. .100 of consensus"
repeat_region   14989. .15308
                /note="AluYb8 repeat: matches 1. .318 of consensus"
misc_feature    15847. .16240
                /note="match: STS: Em:HSPE48H04"
misc_feature    complement(17443. .17932)
                /note="match: GSS: Em:AQ736330"
misc_feature    20566. .20886
                /note="match: GSS: Em:AQ065634"
repeat_region   23685. .23777
                /note="AluJ/FRAM repeat: matches 216. .308 of consensus"
repeat_region   26983. .27024
                /note="21 copies 2 mer tt 78% conserved"
repeat_region   28770. .28887
                /note="AluSq/x repeat: matches 1. .130 of consensus"
repeat_region   30084. .30451
                /note="THE1C repeat: matches 1. .371 of consensus"
repeat_region   35050. .35230
                /note="AluSg/x repeat: matches 128. .309 of consensus"
repeat_region   36151. .36231
                /note="MADE1 repeat: matches 29. .1287 of consensus"
repeat_region   36233. .36546
                /note="L1MC3 repeat: matches 7261. .7573 of consensus"
repeat_region   36846. .36960
                /note="L1MC3 repeat: matches 7147. .7261 of consensus"
repeat_region   38177. .38208
                /note="16 copies 2 mer aa 84% conserved"
misc_feature    complement(39205. .39834)
                /note="match: GSS: Em:AQ590083"
misc_feature    complement(39457. .39820)
                /note="match: GSS: Em:AQ764927"
misc_feature    complement(40213. .40664)
                /note="match: GSS: Em:AQ403675"
repeat_region   40280. .40589
                /note="AluYb8 repeat: matches 1. .318 of consensus"
repeat_region   42125. .42352
                /note="L1ME2 repeat: matches 5917. .6152 of consensus"
repeat_region   42627. .42794
                /note="L1ME2 repeat: matches 5777. .5917 of consensus"
repeat_region   43091. .43208
                /note="L1ME2 repeat: matches 5656. .5777 of consensus"
misc_feature    43969. .44182
                /note="match: GSS: Em:AQ099285"
repeat_region   45387. .45448
                /note="31 copies 2 mer ac 95% conserved"
misc_feature    complement(45763. .46197)
                /note="match: GSS: Em:AQ702452"
repeat_region   45897. .45936
                /note="20 copies 2 mer tg 100% conserved"
repeat_region   46350. .46528
                /note="MSTC repeat: matches 1. .196 of consensus"
repeat_region   50279. .50354
                /note="FRAM/FAM repeat: matches 1. .67 of consensus"
repeat_region   50381. .50519
                /note="LTR37B repeat: matches 330. .466 of consensus"
repeat_region   52901. .52934
                /note="17 copies 2 mer tt 85% conserved"
repeat_region   56672. .56766
                /note="MER3 repeat: matches 4. .98 of consensus"
misc_feature    complement(58791. .59212)
                /note="match: GSS: Em:AQ429967"
misc_feature    62306. .62887
                /note="match: GSS: Em:AQ055709"
repeat_region   66190. .66213
                /note="12 copies 2 mer ta 95% conserved"
repeat_region   66869. .66906
                /note="19 copies 2 mer tt 81% conserved"
misc_feature    complement(67873. .68309)
                /note="match: STS: Em:G52155
                match: GSS: Em:AQ081584"
misc_feature    68717. .69131
                /note="match: GSS: Em:AQ086523"
misc_feature    68721. .68983
                /note="match: STS: Em:HSPE73H6"
misc_feature    complement(69497. .69862)
                /note="match: STS: Em:G38683"
misc_feature    complement(70476. .70778)
                /note="match: GSS: Em:AQ067573"
misc_feature    70903. .71379
                /note="match: GSS: Em:AQ227142"
repeat_region   72915. .73285
                /note="MLT1A1 repeat: matches 1. .365 of consensus"
repeat_region   75116. .75220
                /note="MLT1B repeat: matches 1. .110 of consensus"
gene            75291. .75476
                /gene="dJ733M16.1"
CDS             75291. .75476
                /gene="dJ733M16.1"
                /note="dJ733M16.1 (similar to MT1L (metallothionein 1L))
                match: cDNAs: Em:M26637
                match: ESTs: Em:AI023162
                match: proteins: Sw:O19000 Sw:P14425 Sw:P04355 Sw:P02799
                Tr:Q14823 Sw:P07438 Sw:P04732 Sw:P02795 Sw:P80294
                Sw:P80297 Sw:P02798"
                /pseudo
                /codon_start=1
                /evidence=not_experimental
misc_feature    complement(75394. .75631)
                /note="match: GSS: Em:AQ489791"
misc_feature    75482. .75606
                /note="match: STS: Em:G30412"
repeat_region   76261. .76532
                /note="MLT1B repeat: matches 101. .390 of consensus"
repeat_region   76624. .76876
                /note="MER4A2 repeat: matches 243. .504 of consensus"
repeat_region   77195. .77243
                /note="MER4A2 repeat: matches 196. .243 of consensus"
misc_feature    78694. .79206
                /note="match: GSS: Em:AQ791093"
misc_feature    78732. .79231
                /note="match: GSS: Em:AQ773651"
repeat_region   82221. .82266
                /note="MSTD repeat: matches 352. .393 of consensus"
repeat_region   82572. .82889
                /note="MSTD repeat: matches 1. .352 of consensus"
misc_feature    82598. .83085

  Query Match          44.6%;  Score 260.2;  DB 9;  Length 120825;
  Best Local Similarity   86.4%;  Pred. No. 1.5e-61;
  Matches  312;  Conservative    0;  Mismatches   43;  Indels    6;  Gaps    2;

Qy        221 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 280
              |||||| |||||| |||||||||| |||||||||||||||| |||||| ||||||||||
Db      75239 CCGCTGAGTGTTTGCCTCTTGATCAGGAACTCCTGCTTCTCCCTGCCTCTAAATGGACCC 75298

Qy        281 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 340
              |||||||||||||||||||| ||| ||||| ||||||||||||||||||||||||| |||||
Db      75299 CAACTGCTCCTGCTCGCCTGATGGTTCCTGCGCCTGTGCCGGCTCCTGCAAATGAAAAGA 75358

Qy        341 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAA 400
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75359 GTGCACATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAA 75418

Qy        401 GTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATG 460
              |||||||||||||||||||||||||| | || || || ||||||||||||||||||||
Db      75419 GTGTGCCCAGGGCTGCATCTGCAAAGGGACGTCGGACAACTGCAGCTGCTGTGCCTGATG 75478

Qy        461 TCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGAATTTTTTT 520
               | ||||||||||||||  |||| || | ||||  ||||    || || |     |||||
Db      75479 CCAGGACAGCCCTGCTCTCAGATGTAAATAGAGCAACCTTATAAACCTGG----ATTTTT 75534
```

```
Qy        521 TCCATACAACCCTGCCCCATCTACT--GTATTTTTTTTAATGAAATATGTGAATGATAAT 578
              |  ||||||||||| ||||| | ||   | |||||||  ||||||||||||||||  |||
Db      75535 TTGGTACAACCCTGACCCATTTGCTACATCTTTTTTTCTATGAAATATGTGAATGGCAAT 75594

Qy        579 A 579
              |
Db      75595 A 75595
```

```
RESULT 9
AY077636
LOCUS       AY077636                368 bp    mRNA    linear   PRI 12-FEB-2002
DEFINITION  Homo sapiens zinc-induced metallothionein 2A/1A hybrid mRNA,
            complete cds.
ACCESSION   AY077636
VERSION     AY077636.1  GI:18652805
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 368)
  AUTHORS   Van Weyenbergh,J. and De Ley,M.
  TITLE     Human monocyte zinc-induced mRNA MT2A/1A hybrid
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 368)
  AUTHORS   Van Weyenbergh,J. and De Ley,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JAN-2002) LIMI-CPqGM, FIOCRUZ, Rua Waldemar Falcao
            121, Salvador, BA 40295-001, Brazil
FEATURES             Location/Qualifiers
     source          1. .368
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="ZAP/B2"
                     /cell_type="monocyte"
                     /country="Belgium"
     CDS             14. .199
                     /codon_start=1
                     /product="zinc-induced metallothionein 2A/1A hybrid"
                     /protein_id="AAL76084.1"
                     /db_xref="GI:18652806"
                     /translation="MDPNCSCAAGDSCTCAGSCKCKECKCTSCKKSCCSCCPMSCAKC
                     AQGCICKGASEKCSCCA"
BASE COUNT      106 a     98 c     78 g     86 t
ORIGIN
```

```
  Query Match           44.6%;  Score 259.8;  DB 9;  Length 368;
  Best Local Similarity  89.0%;  Pred. No. 7.6e-62;
  Matches  298;  Conservative    0;  Mismatches   22;  Indels   15;  Gaps    1;

Qy        260 TCCTTGCCTCGAAATGGACCCCAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGC 319
              || |   ||||  ||||| ||||||||||||||| |  | | || ||||||  |||| ||
Db          1 TCTTCAGCTCGCCATGGATCCCAACTGCTCCTGCGCCGCCGGTGACTCCTGCACCTGCGC 60

Qy        320 CGGCTCCTGCAAATGCAAAGAGTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTG 379
              ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db         61 CGGCTCCTGCAAATGCAAAGAGTGCAAATGCACCTCCTGCAAGAAAAGCTGCTGCTCCTG 120

Qy        380 CTGCCCTGTGGGCTGTGCCAAGTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAA 439
              ||||||  || ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTGCCCCATGAGCTGTGCCAAGTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAA 180

Qy        440 GTGCAGCTGCTGTGCCTGATGTCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCT 499
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTGCAGCTGCTGTGCCTGATGTCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCT 240

Qy        500 GCACAAACTTGGAATTTTTTTT--------------CCATACAACCCTGCCCCATCTAC 544
              |||||||||||||||||||||              ||||||||||||| ||||| |||
Db        241 GCACAAACTTGGAATTTTTTTTCCATACAACCCTGACCCATACAACCCTGACCCATTTAC 300

Qy        545 TGTATTTTTTTTAATGAAATATGTGAATGATAATA 579
```

```
              ||||||||||||||||||||||||||||||||||
Db        301 TGTATTTTTTTTAATGAAATATGTGAATGATAATA 335
```

```
RESULT 10
HS1174N9
LOCUS       HS1174N9                72930 bp    DNA     linear   PRI 17-MAY-2000
DEFINITION  Human DNA sequence from clone RP5-1174N9 on chromosome 1p34.1-35.3.
            Contains the gene for a novel protein with IBR domain, a (pseudo?)
            gene for a novel protein similar to MT1E (metallothionein 1E
            (functional)), ESTs, STSs, GSSs and two putative CpG islands,
            complete sequence.
ACCESSION   AL031602
VERSION     AL031602.14  GI:6729581
KEYWORDS    HTG; CpG island; IBR; metallothionein; MT1E.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 72930)
  AUTHORS   Ellington,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jan 21, 2000 this sequence version replaced gi:6634975.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP5-1174N9 is from the library RPCI-5 constructed at the Roswell
            Park Cancer Institute by the group of Pieter de Jong. For further
            details see http://bacpac.med.buffalo.edu/
            VECTOR: pCYPAC2
            This sequence is the entire insert of clone RP5-1174N9.
FEATURES             Location/Qualifiers
     source          1. .72930
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="p34.1-35.3"
                     /clone="RP5-1174N9"
                     /clone_lib="RPCI-5"
     repeat_region   8. .123
                     /note="MIR repeat: matches 66. .181 of consensus"
     repeat_region   124. .418
                     /note="AluSx repeat: matches 1. .297 of consensus"
     repeat_region   419. .477
                     /note="MIR repeat: matches 181. .239 of consensus"
     repeat_region   486. .532
                     /note="Alu repeat: matches 248. .294 of consensus"
     misc_feature    516. .520
                     /note="One clone deleted 5 base pairs."
     repeat_region   536. .652
                     /note="AluSq/x repeat: matches 1. .117 of consensus"
     repeat_region   663. .827
```

```
                /note="MER47 repeat: matches 2130. .2297 of consensus"
repeat_region   1606. .1912
                /note="AluSg repeat: matches 1. .307 of consensus"
repeat_region   2285. .2624
                /note="AluYb8 repeat: matches 2. .312 of consensus"
misc_feature    2872. .3127
                /note="match: GSS: Em:AQ090541"
repeat_region   3112. .3954
                /note="SVA repeat: matches 521. .1382 of consensus"
misc_feature    3278. .6133
                /note="CpG island"
                /evidence=not_experimental
repeat_region   3971. .4355
                /note="SVA repeat: matches 521. .879 of consensus"
repeat_region   4548. .5520
                /note="SVA repeat: matches 1. .954 of consensus"
misc_feature    4757. .4836
                /note="Single clone region. Restriction digest data
                (BAMHI, HINDIII, NCOI) and pUC partners confirm assembly."
repeat_region   5522. .5555
                /note="17 copies 2 mer aa 88% conserved"
repeat_region   6579. .6805
                /note="L2 repeat: matches 2534. .2750 of consensus"
repeat_region   6907. .7064
                /note="FAM repeat: matches -116. .41 of consensus"
repeat_region   7264. .7372
                /note="MIR repeat: matches 77. .196 of consensus"
repeat_region   7887. .8021
                /note="L2 repeat: matches 1950. .2084 of consensus"
repeat_region   8294. .8595
                /note="AluSx repeat: matches 1. .302 of consensus"
repeat_region   8598. .8913
                /note="AluSx repeat: matches 1. .311 of consensus"
repeat_region   8915. .8990
                /note="38 copies 2 mer aa 64% conserved"
repeat_region   9453. .9643
                /note="L1MB3 repeat: matches 5995. .6185 of consensus"
misc_feature    9479. .10051
                /note="match: GSS: Em:AQ590495"
repeat_region   9855. .10167
                /note="AluSx repeat: matches 1. .312 of consensus"
repeat_region   10171. .10566
                /note="L2 repeat: matches 2263. .2679 of consensus"
repeat_region   11172. .11196
                /note="MLT1J repeat: matches 106. .133 of consensus"
repeat_region   11197. .11498
                /note="AluSx repeat: matches 1. .307 of consensus"
repeat_region   11637. .11736
                /note="MLT1I repeat: matches 313. .408 of consensus"
repeat_region   11829. .12036
                /note="L2 repeat: matches 1352. .1567 of consensus"
repeat_region   12077. .12362
                /note="AluJb repeat: matches 1. .286 of consensus"
repeat_region   12363. .12532
                /note="MLT1J repeat: matches 34. .197 of consensus"
repeat_region   12533. .12841
                /note="AluY repeat: matches 1. .309 of consensus"
repeat_region   12842. .13020
                /note="MLT1J repeat: matches 197. .460 of consensus"
repeat_region   13021. .13319
                /note="AluJo repeat: matches 1. .301 of consensus"
repeat_region   13320. .13378
                /note="MLT1J repeat: matches 460. .516 of consensus"
misc_feature    13734. .14018
                /note="match: STS: Em:AA974572"
misc_feature    13976. .15060
                /note="CpG island"
                /evidence=not_experimental
gene            join(14306. .14940,29217. .29422,30628. .30769,32427. .32589,
                33363. .33477,34832. .34972,36532. .36739,40463. .40594,
                41732. .42542)
                /gene="dJ1174N9.1"
mRNA            join(14306. .14940,29217. .29422,30628. .30769,32427. .32589,
                33363. .33477,34832. .34972,36532. .36739,40463. .40594,
                41732. .42542)
                /gene="dJ1174N9.1"
                /product="dJ1174N9.1 (novel protein with IBR domain)"
                /note="match: cDNAs: Em:AF120207 Em:AF120206 Em:X71642
                match: ESTs: Em:AI121963 Em:AW339584 Em:AA413926
                Em:AI152461 Em:R15252 Em:AA647930 Em:AW294203 Em:AA459064
                Em:AA214221 Em:H15240 Em:AA442330 Em:AA214177 Em:AI917452
                Em:AA340392 Em:AW228484 Em:AA064453 Em:AA379493
                Em:AA692791 Em:AA536828 Em:AA906703 Em:AA417244
                Em:AA399232 Em:W27419 Em:R74712 Em:AW105600 Em:C01285
                Em:AI187986 Em:AA992859 Em:AA369389 Em:AA544479
                Em:AA675671 Em:AI952624 Em:AA549628 Em:AW228882
                Em:AW125713 Em:AI029608 Em:AI840373 Em:AA810767
                Em:AA064151 Em:AA194237 Em:T15642 Em:AV275757 Em:T05960
                Em:AA693250 Em:AA193291 Em:AI217132 Em:AA398082
                Em:AA808556 Em:AI220568 Em:AA458874 Em:AI843420
                Em:AA437986 Em:AU045324 Em:AA813317 Em:AI761150 Em:H15178
                Em:T35055 Em:AA781802 Em:N50941"
                /evidence=not_experimental
CDS             join(14306. .14940,29217. .29422,30628. .30769,32427. .32589,
                33363. .33477,34832. .34972,36532. .36739,40463. .40594,
                41732. .42185)
                /gene="dJ1174N9.1"
                /note="supported by FGENESH and GENSCAN
                match: proteins: Tr:P90662 Sw:P50636 Tr:Q9Y4Y1 Tr:P91067
                Sw:P50876 Tr:O76924"
                /codon_start=1
                /evidence=not_experimental
                /product="dJ1174N9.1 (novel protein with IBR domain)"
                /protein_id="CAB92088.1"
                /db_xref="GI:7981255"
                /translation="MGSEKDSESPRSTSLHAAAPDPKCRSGGRRRRLTLHSVFSASAR
                GRRARAKPQAEPPPPAAQPPPAPAPAAAQGPPPEALPAEPAAEAEAEAAAAAAEPGFD
                DEEAAEGGGPGAEEVECPLCLVRLPPERAPRLLSCPHRSCRDCLRHYLRLEISESRVP
                ISCPECSERLNPHDIRLLLADPPLMHKYEEFMLRRYLASDPDCRWCPAPDCGYAVIAY
                GCASCPKLTCEREGCQTEFCYHCKQIWHPNQTCDMARQQRAQTLRVRTKHTSGLSYGQ
                ESGPDDIKPCPRCSAYIIKMNDGSCNHMTCAVCGCEFCWLCMKEISDLHYLSPSGCTF
                WGKKPWSRKKKILWQLGTLIGAPVGISLIAGIAIPAMVIGIPVYYVGRKIHSRYEGRKT
                SKHKRNLAITGGVTLSVIASPVIAAVSVGIGVPIMLAYVYGVVPISLCRGGGCGVSTA
                NGKGVKIEFDEDDGPITVADAWRALKNPSIGESSIEGLTSVLSTSGSPTDGLSVMQGP
                YSETASFAALSGGTLSGGILSSGKGKYSRLEVQADVQKEIFPKDTASLGAISDNASTR
                AMAGSIISSYNPQDRECNNMEIQVDIEAKPSHYQLVSGSSTEDSLHVHAQMAENEEEG
                SGGGGSEEDPPCRHQSCEQKDCLASKPWDISLAQPESIRSDLESSDAQSDDVPDITSD
                ECGSPRSHTAACPSTPRAQGAPSPSAHMNLSALAEGQTVLKPEGGEARV"
misc_feature    14379. .14502
                /gene="dJ1174N9.1"
                /note="Single clone region. Sequence derived from short
                insert library from a single pUC clone. Restriction digest
                data (BAMHI, HINDIII, NCOI) confirm assembly."
repeat_region   15248. .15557
                /note="AluSx repeat: matches 1. .308 of consensus"
repeat_region   16723. .17020
                /note="AluSx repeat: matches 1. .302 of consensus"
repeat_region   17442. .17592
                /note="MIR repeat: matches 34. .189 of consensus"
repeat_region   17967. .18251
                /note="AluJo repeat: matches 26. .291 of consensus"
repeat_region   18339. .18519
                /note="AluJo repeat: matches 120. .300 of consensus"

  Query Match           42.5%;  Score 248;  DB 9;  Length 72930;
  Best Local Similarity  82.7%;  Pred. No. 3.6e-58;
  Matches  296;  Conservative    0;  Mismatches   60;  Indels    2;  Gaps    1;

Qy        224 CTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCCCAA 283
              |||| ||||  ||||      |||| |||| || |||||||  ||||||||||||||||
Db      27463 CTGCTTGTTGGCCTCACCGGTGGGAGCTCCAGCATCTCCTTTGCTCGAAATGGACCCCAA 27522

Qy        284 CTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGAGTG 343
              |||||||||  |  ||| |||||||||    || ||||||||||||| ||||||||||
Db      27523 CTGCTCCTGTGCCACTGGTGGCTCCTGCATGTGCGCCGGCTCCTGCAAGTGCAAAGAGTG 27582
```

```
Qy        344 CAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAAGTG 403
              ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db      27583 CAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCCGTGGGCTGTGCCAAGTG 27642

Qy        404 TGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATGTCC 463
              |||||||||||||||||||||||||| |  |||||||||| ||||| ||||||| ||
Db      27643 TGCCCAGGGCTGCATCTGCAAAGGGGCGTTGGAGAAGTGCAACTGCTATGCCTGACGTGG 27702

Qy        464 GGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGAATTTTTTTTCC 523
              ||||||| ||||||  |||| || | ||||| ||||||||||| |||| |||||
Db      27703 GGACAGCTCTGCTCCCAGATGTAAATAGAGTAACCTGCACAAACCTGGATTTTTTAAAAA 27762

Qy        524 ATACAACCCTGCCCCATCTACTGTATT--TTTTTTAATGAAATATGTGAATGATAATA 579
              ||||||| |||  |||| | ||  |||  |||||  || |||||||| ||||| ||||
Db      27763 ATACAACACTGAGCCATTTGCTTCATTTCTTTTTATATTAAATATGTTAATGACAATA 27820


RESULT 11
BC008408
LOCUS       BC008408                422 bp    mRNA    linear   PRI 12-JUL-2001
DEFINITION  Homo sapiens, metallothionein 1H, clone MGC:14539 IMAGE:4274128,
            mRNA, complete cds.
ACCESSION   BC008408
VERSION     BC008408.1  GI:14250023
KEYWORDS    MGC.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 422)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAY-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:       http://www-shgc.stanford.edu
            Contact:  (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 21 Row: c Column: 18
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 31434.
FEATURES             Location/Qualifiers
     source          1. .422
                     /organism="Homo sapiens"
                     /db_xref="LocusID:4496"
                     /db_xref="taxon:9606"
                     /clone="MGC:14539 IMAGE:4274128"
                     /tissue_type="Prostate"
                     /clone_lib="NIH_MGC_83"
                     /lab_host="DH10B"
                     /note="Vector: pDNR-LIB"
     CDS             72. .257
                     /codon_start=1
                     /product="metallothionein 1H"
                     /protein_id="AAH08408.1"
                     /db_xref="GI:14250024"
                     /translation="MDPNCSCEAGGSCACAGSCKCKKCKCTSCKKSCCSCCPLGCAKC
                     AQGCICKGASEKCSCCA"
BASE COUNT      111 a    116 c     92 g    103 t
ORIGIN

  Query Match           42.3%;  Score 246.4;  DB 9;  Length 422;
  Best Local Similarity   82.0%;  Pred. No. 4.4e-58;
  Matches  297;  Conservative    0;  Mismatches   61;  Indels    4;  Gaps    1;

Qy        221 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 280
              || |||| | || || ||| | | ||||||||| |  || ||| | || | |||||||||
Db         20 CCACTGCCTCTTCTCTTCTCGCTTGGGAACTCCAGTCTCACCTCGGCTTGCAATGGACCC 79

Qy        281 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 340
              |||||||||||||  | ||| |||||||| |||| ||||||||||||||| |||||  |
Db         80 CAACTGCTCCTGCGAGGCTGGTGGCTCCTGCGCCTGCGCCGGCTCCTGCAAGTGCAAGAA 139

Qy        341 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAA 400
              |||||||||||||||||||||||||||||||||||||||| |||||  ||||||||||||
Db        140 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGTTGCCCCCTGGGCTGTGCCAA 199

Qy        401 GTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATG 460
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        200 GTGTGCCCAGGGCTGCATCTGCAAAGGGGCGTCAGAGAAGTGCAGCTGCTGTGCCTGATG 259

Qy        461 TCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGAATTTTTTT 520
              || ||||||||||||    ||||  | | ||| ||||  | | || |  ||| |||||||
Db        260 TCGGGACAGCCCTGCTGTCAGATGAAAACAGAATGACACGTAAAATCCAGGATTTTTTTT 319

Qy        521 TCCATACAACCCTGCCCCATCTACTGTA----TTTTTTTTAATGAAATATGTGAATGATA 576
              |   |||||| | | | ||| | ||  |    |||||||   |||||||||||||| |||
Db        320 TTTCTACAACTCCGACTCATTTGCTACATTCCTTTTTTTCTGTGAAATATGTGAATAATA 379

Qy        577 AT 578
              ||
Db        380 AT 381


RESULT 12
AX330235/c
LOCUS       AX330235                409 bp    DNA     linear   PAT 09-JAN-2002
DEFINITION  Sequence 744 from Patent WO0194629.
ACCESSION   AX330235
VERSION     AX330235.1  GI:18103213
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
            Horrigan,S., Soppet,D.R. and Weaver,Z.
  TITLE     Cancer gene determination and therapeutic screening using signature
            gene sets
  JOURNAL   Patent: WO 0194629-A 744 13-DEC-2001;
            Avalon Pharmaceuticals (US)
FEATURES             Location/Qualifiers
     source          1. .409
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      116 a     93 c    109 g     87 t      4 others
ORIGIN

  Query Match           42.1%;  Score 245.2;  DB 6;  Length 409;
  Best Local Similarity   88.0%;  Pred. No. 9.6e-58;
  Matches  324;  Conservative    0;  Mismatches   35;  Indels    9;  Gaps    5;

Qy        221 CCGCTGCGTGTTTTCCTCTTG-ATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACC 279
              |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db        382 CCGCTGCGTGTTTTCCTCTTGAATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACC 323

Qy        280 CCAACTGCTCCTGC-TCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAA 338
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        322 CCAACTGCTCCTGCNTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAA 263

Qy        339 GAGTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCC-TGCTGCCCTGTGGGCTGTGC 397
              |||||||||||||||||||||||||||||||||||||| ||| ||||||||||||||
```

```
Db        262 GAGTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCNTGCNTNCCTGTGGGCTGTGC 203

Qy        398 CAAGTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTG 457
              ||||||||||||||||||||||||||||||| | ||||| ||||||||||||||||||||
Db        202 CAAGTGTGCCCAGGGCTGCATCTGCAAAGGGACGTCAGACAAGTGCAGCTGCTGTGCCTG 143

Qy        458 ATGTCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGA----A 513
              ||| | |||||||  |||||  |||| || | ||||  ||||  | |||| ||||
Db        142 ATGCCAGGACAGCTGTGCTCTCAGATGTAAATAGAGCAACCTATATAAACCTGGATTTTT 83

Qy        514 TTTTTTTTCCATACAACCCTGCCCCATCTACT--GTATTTTTTTTAATGAAATATGTGAA 571
              ||||||||   |||||||||| ||| | | ||   | |||||||  ||||||||||||||
Db         82 TTTTTTTTTTGTACAACCCTGACCCGTTTGCTACATCTTTTTTTCTATGAAATATGTGAA 23

Qy        572 TGATAATA 579
              ||  ||||
Db         22 TGGCAATA 15


RESULT 13
AX331811/c
LOCUS       AX331811                409 bp    DNA     linear   PAT 09-JAN-2002
DEFINITION  Sequence 2320 from Patent WO0194629.
ACCESSION   AX331811
VERSION     AX331811.1  GI:18122445
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
            Horrigan,S., Soppet,D.R. and Weaver,Z.
  TITLE     Cancer gene determination and therapeutic screening using signature
            gene sets
  JOURNAL   Patent: WO 0194629-A 2320 13-DEC-2001;
            Avalon Pharmaceuticals (US)
FEATURES             Location/Qualifiers
     source          1. .409
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      116 a     93 c    109 g     87 t      4 others
ORIGIN

  Query Match           42.1%;  Score 245.2;  DB 6;  Length 409;
  Best Local Similarity  88.0%;  Pred. No. 9.6e-58;
  Matches  324;  Conservative    0;  Mismatches   35;  Indels    9;  Gaps    5;

Qy        221 CCGCTGCGTGTTTTCCTCTTG-ATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACC 279
              ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        382 CCGCTGCGTGTTTTCCTCTTGAATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACC 323

Qy        280 CCAACTGCTCCTGC-TCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAA 338
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        322 CCAACTGCTCCTGCNTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAA 263

Qy        339 GAGTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCC-TGCTGCCCTGTGGGCTGTGC 397
              ||||||||||||||||||||||||||||||||||||||| |||   ||||||||||||||
Db        262 GAGTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCNTGCNTNCCTGTGGGCTGTGC 203

Qy        398 CAAGTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTG 457
              ||||||||||||||||||||||||||||||| | ||||| ||||||||||||||||||||
Db        202 CAAGTGTGCCCAGGGCTGCATCTGCAAAGGGACGTCAGACAAGTGCAGCTGCTGTGCCTG 143

Qy        458 ATGTCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGA----A 513
              ||| | |||||||  |||||  |||| || | ||||  ||||  | |||| ||||
Db        142 ATGCCAGGACAGCTGTGCTCTCAGATGTAAATAGAGCAACCTATATAAACCTGGATTTTT 83

Qy        514 TTTTTTTTCCATACAACCCTGCCCCATCTACT--GTATTTTTTTTAATGAAATATGTGAA 571
              ||||||||   |||||||||| ||| | | ||   | |||||||  ||||||||||||||
Db         82 TTTTTTTTTTGTACAACCCTGACCCGTTTGCTACATCTTTTTTTCTATGAAATATGTGAA 23

Qy        572 TGATAATA 579
              ||  ||||
Db         22 TGGCAATA 15


RESULT 14
AX336059/c
LOCUS       AX336059                409 bp    DNA     linear   PAT 09-JAN-2002
DEFINITION  Sequence 6568 from Patent WO0194629.
ACCESSION   AX336059
VERSION     AX336059.1  GI:18126778
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
            Horrigan,S., Soppet,D.R. and Weaver,Z.
  TITLE     Cancer gene determination and therapeutic screening using signature
            gene sets
  JOURNAL   Patent: WO 0194629-A 6568 13-DEC-2001;
            Avalon Pharmaceuticals (US)
FEATURES             Location/Qualifiers
     source          1. .409
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      116 a     93 c    109 g     87 t      4 others
ORIGIN

  Query Match           42.1%;  Score 245.2;  DB 6;  Length 409;
  Best Local Similarity  88.0%;  Pred. No. 9.6e-58;
  Matches  324;  Conservative    0;  Mismatches   35;  Indels    9;  Gaps    5;

Qy        221 CCGCTGCGTGTTTTCCTCTTG-ATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACC 279
              ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        382 CCGCTGCGTGTTTTCCTCTTGAATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACC 323

Qy        280 CCAACTGCTCCTGC-TCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAA 338
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        322 CCAACTGCTCCTGCNTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAA 263

Qy        339 GAGTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCC-TGCTGCCCTGTGGGCTGTGC 397
              ||||||||||||||||||||||||||||||||||||||| |||   ||||||||||||||
Db        262 GAGTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCNTGCNTNCCTGTGGGCTGTGC 203

Qy        398 CAAGTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTG 457
              ||||||||||||||||||||||||||||||| | ||||| ||||||||||||||||||||
Db        202 CAAGTGTGCCCAGGGCTGCATCTGCAAAGGGACGTCAGACAAGTGCAGCTGCTGTGCCTG 143

Qy        458 ATGTCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGA----A 513
              ||| | |||||||  |||||  |||| || | ||||  ||||  | |||| ||||
Db        142 ATGCCAGGACAGCTGTGCTCTCAGATGTAAATAGAGCAACCTATATAAACCTGGATTTTT 83

Qy        514 TTTTTTTTCCATACAACCCTGCCCCATCTACT--GTATTTTTTTTAATGAAATATGTGAA 571
              ||||||||   |||||||||| ||| | | ||   | |||||||  ||||||||||||||
Db         82 TTTTTTTTTTGTACAACCCTGACCCGTTTGCTACATCTTTTTTTCTATGAAATATGTGAA 23

Qy        572 TGATAATA 579
              ||  ||||
Db         22 TGGCAATA 15


RESULT 15
AX336280/c
LOCUS       AX336280                409 bp    DNA     linear   PAT 09-JAN-2002
DEFINITION  Sequence 6789 from Patent WO0194629.
ACCESSION   AX336280
VERSION     AX336280.1  GI:18126999
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE   1
  AUTHORS   Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
            Horrigan,S., Soppet,D.R. and Weaver,Z.
  TITLE     Cancer gene determination and therapeutic screening using signature
            gene sets
  JOURNAL   Patent: WO 0194629-A 6789 13-DEC-2001;
            Avalon Pharmaceuticals (US)
FEATURES             Location/Qualifiers
     source          1. .409
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      116 a     93 c    109 g     87 t      4 others
ORIGIN

  Query Match           42.1%;  Score 245.2;  DB 6;  Length 409;
  Best Local Similarity 88.0%;  Pred. No. 9.6e-58;
  Matches  324;  Conservative    0;  Mismatches   35;  Indels    9;  Gaps    5;

Qy        221 CCGCTGCGTGTTTTCCTCTTG-ATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACC 279
              |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db        382 CCGCTGCGTGTTTTCCTCTTGAATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACC 323

Qy        280 CCAACTGCTCCTGC-TCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAA 338
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        322 CCAACTGCTCCTGCNTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAA 263

Qy        339 GAGTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCC-TGCTGCCCTGTGGGCTGTGC 397
              |||||||||||||||||||||||||||||||||||||| |||   |||||||||||||
Db        262 GAGTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCNTGCNTNCCTGTGGGCTGTGC 203

Qy        398 CAAGTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTG 457
              |||||||||||||||||||||||||||||| | ||||| ||||||||||||||||||||
Db        202 CAAGTGTGCCCAGGGCTGCATCTGCAAAGGGACGTCAGACAAGTGCAGCTGCTGTGCCTG 143

Qy        458 ATGTCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGA----A 513
              ||| | |||||||  |||||  |||| || | ||||  ||||  | |||| ||||
Db        142 ATGCCAGGACAGCTGTGCTCTCAGATGTAAATAGAGCAACCTATATAAACCTGGATTTTT 83

Qy        514 TTTTTTTTCCATACAACCCTGCCCCATCTACT--GTATTTTTTTTAATGAAATATGTGAA 571
              ||||||||   |||||||||| ||| | | ||   | |||||||  ||||||||||||||
Db         82 TTTTTTTTTTGTACAACCCTGACCCGTTTGCTACATCTTTTTTTCTATGAAATATGTGAA 23

Qy        572 TGATAATA 579
              ||  ||||
Db         22 TGGCAATA 15


Search completed: July 16, 2003, 13:27:55
Job time : 1063.5 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 16, 2003, 03:41:41 ; Search time 99.5867 Seconds
(without alignments)
13183.634 Million cell updates/sec

Title: US-09-981-353-114
Perfect score: 583
Sequence: 1 ccctgtaggtgctgtccttg.....tatgtgaatgataatagtca 583

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	277.4	47.6	415	24	ABN97299	Gene #3797 used to
	2	277.4	47.6	415	24	ABK64825	Human benign prost
	3	277.4	47.6	415	24	ABL62482	Colon adenocarcino
	4	277.4	47.6	415	24	ABL68578	Kidney cancer rela
	5	270.6	46.4	448	22	AAI93455	Human polynucleoti
	6	248	42.5	573	22	AAI61205	Human polynucleoti
	7	247	42.4	446	22	AAI59419	Human polynucleoti
	8	247	42.4	601	22	AAK52094	Human polynucleoti
c	9	245.2	42.1	409	24	ABN96205	Gene #2703 used to
c	10	245.2	42.1	409	24	ABK64607	Human benign prost
c	11	245.2	42.1	409	24	ABL62407	Colon adenocarcino
c	12	245.2	42.1	409	24	ABL63983	Breast cancer rela
c	13	245.2	42.1	409	24	ABL68231	Kidney cancer rela
c	14	245.2	42.1	409	24	ABL68452	Kidney cancer rela
c	15	245.2	42.1	409	24	ABL68852	Kidney cancer rela
	16	232.8	39.9	453	21	AAF16386	Human prostate can
	17	227.6	39.0	682	21	AAF16055	Human prostate can
	18	227.2	39.0	420	22	AAI93800	Human polynucleoti
	19	223.6	38.4	389	22	AAS45033	cDNA encoding nove
	20	221.8	38.0	641	22	AAS23089	DNA encoding novel
	21	219.8	37.7	363	16	AAT20036	Human gene signatu
	22	219.8	37.7	367	24	ABN97265	Gene #3763 used to
	23	219.8	37.7	367	24	ABK64818	Human benign prost
	24	219.8	37.7	367	24	ABL68636	Kidney cancer rela
	25	217.4	37.3	353	16	AAT20977	Human gene signatu
	26	216.8	37.2	557	22	AAK51621	Human polynucleoti
	27	216.4	37.1	500	21	AAC99226	Human pancreatic c
	28	216	37.0	646	22	AAF44864	Human breast cance
c	29	216	37.0	673	22	AAH30041	TRO005 kappa chain
c	30	216	37.0	678	22	AAH41664	Human interleukin
c	31	216	37.0	678	22	AAH41667	Human interleukin
c	32	216	37.0	678	22	AAH41670	Human interleukin
c	33	216	37.0	678	22	AAH30010	Anti-IL8 monoclona
c	34	216	37.0	678	22	AAH30013	Anti-IL8 monoclona
c	35	216	37.0	678	22	AAH30016	Anti-IL8 monoclona
c	36	216	37.0	923	21	AAA95785	Human immune syste
c	37	216	37.0	944	22	AAF44892	Human breast cance
c	38	216	37.0	948	24	AAS99475	Anti-human AILIM m
c	39	216	37.0	958	22	AAC66520	Human immune syste
c	40	216	37.0	970	24	AAS99477	Anti-human AILIM m
c	41	214.4	36.8	645	16	AAQ67896	cDNA contg. an ORF
c	42	214.4	36.8	678	22	AAH41652	Human interleukin
c	43	214.4	36.8	678	22	AAH29998	Anti-IL8 monoclona
c	44	214.4	36.8	726	17	AAT18060	Monoclonal antibod
c	45	214.4	36.8	758	20	AAX01213	Human antiFc epsil

ALIGNMENTS

RESULT 1
ABN97299
ID ABN97299 standard; DNA; 415 BP.
XX
AC ABN97299;
XX
DT 13-AUG-2002 (first entry)
XX
DE Gene #3797 used to diagnose liver cancer.
XX
KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW metastatic liver tumour; cytostatic; expression profile; disease state;
KW disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
OS Homo sapiens.
XX
PN WO200229103-A2.
XX
PD 11-APR-2002.
XX
PF 02-OCT-2001; 2001WO-US30589.
XX
PR 02-OCT-2000; 2000US-237054P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX
DR WPI; 2002-426119/45.
XX
PT Diagnosing and detecting the progression of liver cancer,
PT hepatocellular carcinoma or metastatic liver tumor in a patient,

PT involves detecting the level of expression of two or more genes in a
PT liver tissue sample -
XX
PS Claim 1; SEQ ID NO 3797; 298pp; English.
XX
CC The invention relates to a novel method for diagnosing and detecting the
CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC tumour in a patient, and differentiating metastatic liver cancer from
CC hepatocellular carcinoma in a patient, involving detecting the level of
CC expression of two or more genes represented in ABN93503-ABN97455 in a
CC tissue sample. The method of the invention has hepatotropic, and
CC cytostatic activity. The method is useful for diagnosing and detecting
CC the progression of liver cancer, hepatocellular carcinoma and metastatic
CC liver carcinoma in a patient. The method is useful for identifying
CC expression profiles which serve as useful diagnostic markers as well as
CC markers that can be used to monitor disease states, disease progression,
CC drug toxicity, drug efficacy and drug metabolism.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 415 BP; 110 A; 102 C; 87 G; 115 T; 1 other;

Query Match 47.6%; Score 277.4; DB 24; Length 415;
Best Local Similarity 88.3%; Pred. No. 1.2e-68;
Matches 326; Conservative 1; Mismatches 32; Indels 10; Gaps 2;

```
Qy        221 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 60

Qy        281 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 120

Qy        341 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAA 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:||
Db        121 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCMAA 180

Qy        401 GTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATG 460
              |||||||||||||||||||||||||| | ||||| ||||||||||||||||||||||||
Db        181 GTGTGCCCAGGGCTGCATCTGCAAAGGGACGTCAGACAAGTGCAGCTGCTGTGCCTGATG 240

Qy        461 TCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGA-------- 512
              | ||||||| |||||  |||| || | ||||  ||||  | |||| ||||
Db        241 CCAGGACAGCTGTGCTCTCAGATGTAAATAGAGCAACCTATATAAACCTGGATTTTTTTT 300

Qy        513 ATTTTTTTTCCATACAACCCTGCCCCATCTACT--GTATTTTTTTTAATGAAATATGTGA 570
              ||||||||   |||||||||| ||| | | ||   | |||||||  ||||||||||||
Db        301 TTTTTTTTTTTGTACAACCCTGACCCGTTTGCTACATCTTTTTTTCTATGAAATATGTGA 360

Qy        571 ATGATAATA 579
              |||  ||||
Db        361 ATGGCAATA 369
```

RESULT 2
ABK64825
ID ABK64825 standard; DNA; 415 BP.
XX
AC ABK64825;
XX
DT 18-JUN-2002 (first entry)
XX
DE Human benign prostatic hyperplasia gene #720.
XX
KW Human; benign prostatic hyperplasia; BPH; prostate cancer; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200212440-A2.
XX
PD 14-FEB-2002.
XX
PF 07-AUG-2001; 2001WO-US24708.
XX
PR 07-AUG-2000; 2000US-223323P.
PR 05-JUN-2001; 2001US-0873319.
XX
PA (GENE-) GENE LOGIC INC.
PA (NISB) JAPAN TOBACCO INC.
XX
PI Munger WE, Kulkarni P, Getzenberg RH, Waga I, Yamamoto J;
XX
DR WPI; 2002-257476/30.
XX
PT Identifying drugs for and diagnosing benign prostatic hyperplasia, by
PT detecting expression levels of one or more genes in prostate cells from
PT patient that are differentially regulated compared to normal prostate
PT cells -
XX
PS Disclosure; Page 402; 444pp; English.
XX
CC The invention relates to a method of diagnosing (I) the onset or
CC progression of benign prostatic hyperplasia (BPH), or screening (II) for
CC or identifying an agent that modulates the onset or progression of BPH.
CC The method is based on changes in gene expression in BPH tissue isolated
CC from patients exhibiting different clinical states of prostate
CC hyperplasia as compared to normal prostate tissue. (I) comprises
CC detecting the expression levels of one or more genes in prostate cells
CC from the subject that are differentially regulated compared to normal
CC prostate cells. (II) comprises preparing a first gene expression profile
CC of BPH cells or BPH-like cell population, exposing the cells to the
CC agent, preparing a second gene expression profile of the agent exposed
CC cells, and comparing the first and second gene expression profiles.
CC (I) is useful for diagnosing the onset or progression of BPH. (II) is
CC useful for identifying an agent that modulates the onset or progression
CC of BPH. The methods are useful to present information identifying
CC the expression level in a tissue or cells, by comparing the expression
CC level of genes given in the specification in the tissue or cells to the
CC level of expression of gene in the database, and displaying the
CC expression levels of at least one gene in the tissue or cell sample
CC compared to the expression level in BPH. Agents using (II) are useful for
CC treating BPH or prostate cancer. ABK64106-ABK64860 represent human
CC benign prostatic hyperplasia gene sequences of the invention.
XX
SQ Sequence 415 BP; 110 A; 102 C; 87 G; 115 T; 1 other;

Query Match 47.6%; Score 277.4; DB 24; Length 415;
Best Local Similarity 88.3%; Pred. No. 1.2e-68;
Matches 326; Conservative 1; Mismatches 32; Indels 10; Gaps 2;

```
Qy        221 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 60

Qy        281 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 120

Qy        341 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAA 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:||
Db        121 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCMAA 180

Qy        401 GTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATG 460
              |||||||||||||||||||||||||| | ||||| ||||||||||||||||||||||||
Db        181 GTGTGCCCAGGGCTGCATCTGCAAAGGGACGTCAGACAAGTGCAGCTGCTGTGCCTGATG 240

Qy        461 TCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGA-------- 512
              | ||||||| |||||  |||| || | ||||  ||||  | |||| ||||
Db        241 CCAGGACAGCTGTGCTCTCAGATGTAAATAGAGCAACCTATATAAACCTGGATTTTTTTT 300

Qy        513 ATTTTTTTTCCATACAACCCTGCCCCATCTACT--GTATTTTTTTTAATGAAATATGTGA 570
              ||||||||   |||||||||| ||| | | ||   | |||||||  ||||||||||||
Db        301 TTTTTTTTTTTGTACAACCCTGACCCGTTTGCTACATCTTTTTTTCTATGAAATATGTGA 360
```

```
Qy        571 ATGATAATA 579
              |||  ||||
Db        361 ATGGCAATA 369
```

RESULT 3
ABL62482
ID ABL62482 standard; DNA; 415 BP.
XX
AC ABL62482;
XX
DT 15-MAY-2002 (first entry)
XX
DE Colon adenocarcinoma related gene sequence SEQ ID NO:819.
XX
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200194629-A2.
XX
PD 13-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US10838.
XX
PR 05-JUN-2000; 2000US-209473P.
PR 05-JUN-2000; 2000US-209531P.
PR 18-SEP-2000; 2000US-233133P.
PR 18-SEP-2000; 2000US-233617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 20-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.
PR 25-SEP-2000; 2000US-234923P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-235077P.
PR 25-SEP-2000; 2000US-235082P.
PR 25-SEP-2000; 2000US-235134P.
PR 25-SEP-2000; 2000US-235280P.
PR 26-SEP-2000; 2000US-235637P.
PR 26-SEP-2000; 2000US-235638P.
PR 27-SEP-2000; 2000US-235711P.
PR 27-SEP-2000; 2000US-235720P.
PR 27-SEP-2000; 2000US-235840P.
PR 27-SEP-2000; 2000US-235863P.
PR 28-SEP-2000; 2000US-236028P.
PR 28-SEP-2000; 2000US-236032P.
PR 28-SEP-2000; 2000US-236033P.
PR 28-SEP-2000; 2000US-236034P.
PR 28-SEP-2000; 2000US-236109P.
PR 28-SEP-2000; 2000US-236111P.
PR 29-SEP-2000; 2000US-236842P.
PR 29-SEP-2000; 2000US-236891P.
PR 02-OCT-2000; 2000US-237172P.
PR 02-OCT-2000; 2000US-237173P.
PR 02-OCT-2000; 2000US-237278P.
PR 02-OCT-2000; 2000US-237294P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237316P.
PR 03-OCT-2000; 2000US-237425P.
PR 03-OCT-2000; 2000US-237598P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.
PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.
XX
PA (AVAL-) AVALON PHARM.
XX
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX
DR WPI; 2002-188264/24.
XX
PT Screening for anti-neoplastic agent involves exposing cells to a
PT chemical agent to be tested for anti-neoplastic activity, and
PT determining a change in expression of a gene of a signature gene set -
XX
PS Claim 1; SEQ ID 819; 44pp; English.
XX
CC The present invention describes a method (M1) for screening for an
CC anti-neoplastic agent. The method involves exposing cells to a chemical
CC agent to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening
CC an anti-neoplastic agent, and can be used for producing a product which
CC is the data collected with respect to the anti-neoplastic agent as a
CC result of M1, and the data is sufficient to convey the chemical
CC structure and/or properties of the agent. M1 can be used in the
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC carcinoma, papillary carcinoma and Wilm's tumour.
XX
SQ Sequence 415 BP; 110 A; 102 C; 87 G; 115 T; 1 other;

Query Match 47.6%; Score 277.4; DB 24; Length 415;
Best Local Similarity 88.3%; Pred. No. 1.2e-68;
Matches 326; Conservative 1; Mismatches 32; Indels 10; Gaps 2;

```
Qy        221 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 60

Qy        281 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 120

Qy        341 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAA 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:||
Db        121 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCMAA 180

Qy        401 GTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATG 460
              |||||||||||||||||||||||||| | ||||| ||||||||||||||||||||||||
Db        181 GTGTGCCCAGGGCTGCATCTGCAAAGGGACGTCAGACAAGTGCAGCTGCTGTGCCTGATG 240

Qy        461 TCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGA-------- 512
              | ||||||| ||||| |||| || | |||  |||| | |||| ||||
Db        241 CCAGGACAGCTGTGCTCTCAGATGTAAATAGAGCAACCTATATAAACCTGGATTTTTTTT 300

Qy        513 ATTTTTTTTCCATACAACCCTGCCCCATCTACT--GTATTTTTTTTAATGAAATATGTGA 570
              ||||||||  |||||||||| ||| | | ||   | |||||||  ||||||||||||
Db        301 TTTTTTTTTTTGTACAACCCTGACCCGTTTGCTACATCTTTTTTTCTATGAAATATGTGA 360

Qy        571 ATGATAATA 579
              |||  ||||
Db        361 ATGGCAATA 369
```

RESULT 4
ABL68578
ID ABL68578 standard; DNA; 415 BP.
XX
AC ABL68578;
XX
DT 15-MAY-2002 (first entry)
XX
DE Kidney cancer related gene sequence SEQ ID NO:6915.

```
XX
KW   Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW   stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW   cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200194629-A2.
XX
PD   13-DEC-2001.
XX
PF   30-MAY-2001; 2001WO-US10838.
XX
PR   05-JUN-2000; 2000US-209473P.
PR   05-JUN-2000; 2000US-209531P.
PR   18-SEP-2000; 2000US-233133P.
PR   18-SEP-2000; 2000US-233617P.
PR   20-SEP-2000; 2000US-234009P.
PR   20-SEP-2000; 2000US-234034P.
PR   20-SEP-2000; 2000US-234052P.
PR   22-SEP-2000; 2000US-234509P.
PR   22-SEP-2000; 2000US-234567P.
PR   25-SEP-2000; 2000US-234923P.
PR   25-SEP-2000; 2000US-234924P.
PR   25-SEP-2000; 2000US-235077P.
PR   25-SEP-2000; 2000US-235082P.
PR   25-SEP-2000; 2000US-235134P.
PR   25-SEP-2000; 2000US-235280P.
PR   26-SEP-2000; 2000US-235637P.
PR   26-SEP-2000; 2000US-235638P.
PR   27-SEP-2000; 2000US-235711P.
PR   27-SEP-2000; 2000US-235720P.
PR   27-SEP-2000; 2000US-235840P.
PR   27-SEP-2000; 2000US-235863P.
PR   28-SEP-2000; 2000US-236028P.
PR   28-SEP-2000; 2000US-236032P.
PR   28-SEP-2000; 2000US-236033P.
PR   28-SEP-2000; 2000US-236034P.
PR   28-SEP-2000; 2000US-236109P.
PR   28-SEP-2000; 2000US-236111P.
PR   29-SEP-2000; 2000US-236842P.
PR   29-SEP-2000; 2000US-236891P.
PR   02-OCT-2000; 2000US-237172P.
PR   02-OCT-2000; 2000US-237173P.
PR   02-OCT-2000; 2000US-237278P.
PR   02-OCT-2000; 2000US-237294P.
PR   02-OCT-2000; 2000US-237295P.
PR   02-OCT-2000; 2000US-237316P.
PR   03-OCT-2000; 2000US-237425P.
PR   03-OCT-2000; 2000US-237598P.
PR   03-OCT-2000; 2000US-237604P.
PR   03-OCT-2000; 2000US-237606P.
PR   03-OCT-2000; 2000US-237608P.
PR   01-NOV-2000; 2000US-244867P.
PR   01-NOV-2000; 2000US-245084P.
XX
PA   (AVAL-) AVALON PHARM.
XX
PI   Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI   Soppet DR, Weaver Z;
XX
DR   WPI; 2002-188264/24.
XX
PT   Screening for anti-neoplastic agent involves exposing cells to a
PT   chemical agent to be tested for anti-neoplastic activity, and
PT   determining a change in expression of a gene of a signature gene set
XX
PS   Claim 1; SEQ ID 6915; 44pp; English.
XX
CC   The present invention describes a method (M1) for screening for an
CC   anti-neoplastic agent. The method involves exposing cells to a chemical
CC   agent to be tested for anti-neoplastic activity, determining a change in
CC   expression of at least one gene (I) of a signature gene set, where (I)
CC   comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC   to ABL70110), or is at least 95% identical to (S), where a change in
CC   expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC   activity and can be used in gene therapy. M1 can be used for screening
CC   an anti-neoplastic agent, and can be used for producing a product which
CC   is the data collected with respect to the anti-neoplastic agent as a
CC   result of M1, and the data is sufficient to convey the chemical
CC   structure and/or properties of the agent. M1 can be used in the
CC   treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC   oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC   adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC   infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC   carcinoma, papillary carcinoma and Wilm's tumour.
XX
SQ   Sequence 415 BP; 110 A; 102 C; 87 G; 115 T; 1 other;

  Query Match             47.6%;  Score 277.4;  DB 24;  Length 415;
  Best Local Similarity   88.3%;  Pred. No. 1.2e-68;
  Matches  326;  Conservative    1;  Mismatches   32;  Indels   10;  Gaps    2;

Qy        221 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 60

Qy        281 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 120

Qy        341 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAA 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:||
Db        121 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCMAA 180

Qy        401 GTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATG 460
              |||||||||||||||||||||||||||| | ||||| ||||||||||||||||||||||
Db        181 GTGTGCCCAGGGCTGCATCTGCAAAGGGACGTCAGACAAGTGCAGCTGCTGTGCCTGATG 240

Qy        461 TCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGA-------- 512
               | ||||||| |||||  |||| || | ||||  |||| | |||| ||||
Db        241 CCAGGACAGCTGTGCTCTCAGATGTAAATAGAGCAACCTATATAAACCTGGATTTTTTTT 300

Qy        513 ATTTTTTTTCCATACAACCCTGCCCCATCTACT--GTATTTTTTTTAATGAAATATGTGA 570
              ||||||||   ||||||||| ||| | | ||   | |||||||  ||||||||||||
Db        301 TTTTTTTTTTTGTACAACCCTGACCCGTTTGCTACATCTTTTTTTCTATGAAATATGTGA 360

Qy        571 ATGATAATA 579
              |||  ||||
Db        361 ATGGCAATA 369


RESULT 5
AAI93455
ID   AAI93455 standard; cDNA; 448 BP.
XX
AC   AAI93455;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 13515.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation; ss.
XX
OS   Homo sapiens.
XX
PN   WO200164835-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US04927.
XX
```

```
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-514838/56.
DR   P-PSDB; AAO13524.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing
PT   diagnosing and treating e.g. leukaemia, inflammation and immune
PT   disorders -
XX
PS   Claim 1; SEQ ID NO 13515; 1399pp + Sequence Listing; English.
XX
CC   The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC   the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 448 BP; 94 A; 116 C; 119 G; 119 T; 0 other;

  Query Match           46.4%;  Score 270.6;  DB 22;  Length 448;
  Best Local Similarity  93.7%;  Pred. No. 1e-66;
  Matches  282;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy        221 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         54 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 113

Qy        281 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        114 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 173

Qy        341 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAA 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        174 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAA 233

Qy        401 GTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATG 460
              |||||||||||||||||||||||||||| | ||||| ||||||||||||||||||||||||
Db        234 GTGTGCCCAGGGCTGCATCTGCAAAGGGACGTCAGACAAGTGCAGCTGCTGTGCCTGATG 293

Qy        461 TCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGAATTTTTTT 520
              | |||||||  |||||  |||| || | ||||  ||||  | |||| |||| |||||||
Db        294 CCAGGACAGCTGTGCTCTCAGATGTAAATAGAGCAACCTATATAAACCTGGATTTTTTTT 353

Qy        521 T 521
              |
Db        354 T 354


RESULT 6
AAI61205
ID   AAI61205 standard; cDNA; 573 BP.
XX
AC   AAI61205;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 5194.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia; ss.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0653450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C,  Xue AJ,  Yang Y,  Zhang J;
PI   Zhao QA,  Zhou P,  Goodrich R,  Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   P-PSDB; AAM42049.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
PS   Claim 1; SEQ ID NO 5194; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence 573 BP; 132 A; 171 C; 141 G; 129 T; 0 other;

  Query Match           42.5%;  Score 248;  DB 22;  Length 573;
  Best Local Similarity  82.3%;  Pred. No. 2.9e-60;
  Matches  298;  Conservative    0;  Mismatches   60;  Indels    4;  Gaps    1;

Qy        221 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 280
              || |||| | || || ||| | | |||||||| |  || ||| | || | |||||||||
Db        186 CCACTGCCTCTTCTCTTCTCGCTTGGGAACTCCAGTCTCACCTCGGCTTGCAATGGACCC 245

Qy        281 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 340
              |||||||||||||  | ||| |||||||| ||||| |||||||||||||| |||||| |
Db        246 CAACTGCTCCTGCGAGGCTGGTGGCTCCTGCGCCTGCGCCGGCTCCTGCAAGTGCAAAAA 305

Qy        341 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAA 400
              |||||||||||||||||||||||||||||||||||||||| |||||  ||||||||||||
Db        306 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGTTGCCCCCTGGGCTGTGCCAA 365

Qy        401 GTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATG 460
```

```
                ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        366 GTGTGCCCAGGGCTGCATCTGCAAAGGGGCGTCAGAGAAGTGCAGCTGCTGTGCCTGATG 425

Qy        461 TCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGAATTTTTTT 520
                || ||||||||||||    ||||  | | ||| ||||  | | || |  ||| |||||||
Db        426 TCGGGACAGCCCTGCTGTCAGATGAAAACAGAATGACACGTAAAATCCAGGATTTTTTTT 485

Qy        521 TCCATACAACCCTGCCCCCATCTACTGTA----TTTTTTTTAATGAAATATGTGAATGATA 576
                |   |||||| | | | ||| | ||  |    |||||||   |||||||||||||| |||
Db        486 TTTCTACAACTCCGACTCATTTGCTACATTCCTTTTTTTCTGTGAAATATGTGAATAATA 545

Qy        577 AT 578
                ||
Db        546 AT 547
```

RESULT 7
AAI59419
ID AAI59419 standard; cDNA; 446 BP.
XX
AC AAI59419;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 1622.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR P-PSDB; AAM40263.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Claim 1; SEQ ID NO 1622; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 446 BP; 108 A; 119 C; 105 G; 114 T; 0 other;

 Query Match 42.4%; Score 247; DB 22; Length 446;
 Best Local Similarity 82.1%; Pred. No. 5e-60;
 Matches 298; Conservative 0; Mismatches 60; Indels 5; Gaps 1;

```
Qy        221 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 280
                || |||| | || || ||| | | ||||||||| |  || ||| | || | |||||||||
Db         56 CCACTGCCTCTTCTCTTCTCGCTTGGGAACTCCAGTCTCACCTCGGCTTGCAATGGACCC 115

Qy        281 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 340
                |||||||||||||  | ||| ||||||||| ||||| |||||||||||||| |||||| |
Db        116 CAACTGCTCCTGCGAGGCTGGTGGCTCCTGCGCCTGCGCCGGCTCCTGCAAGTGCAAAAA 175

Qy        341 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAA 400
                |||||||||||||||||||||||||||||||||||||||| |||||  ||||||||||||
Db        176 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGTTGCCCCCTGGGCTGTGCCAA 235

Qy        401 GTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATG 460
                |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        236 GTGTGCCCAGGGCTGCATCTGCAAAGGGGCGTCAGAGAAGTGCAGCTGCTGTGCCTGATG 295

Qy        461 TCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGAATTTTTTT 520
                || ||||||||||||    ||||  | | ||| ||||  | | || |  ||| |||||||
Db        296 TCGGGACAGCCCTGCTGTCAGATGAAAACAGAATGACACGTAAAATCCAGGATTTTTTTT 355

Qy        521 TCCATACAACCCTGCCCCATCTACTGTA-----TTTTTTTTAATGAAATATGTGAATGAT 575
                |   |||||| | | | ||| | ||  |     |||||||   |||||||||||||| ||
Db        356 TTTCTACAACTCCGACTCATTTGCTACATTCCTTTTTTTTCTGTGAAATATGTGAATAAT 415

Qy        576 AAT 578
                |||
Db        416 AAT 418
```

RESULT 8
AAK52094
ID AAK52094 standard; cDNA; 601 BP.
XX
AC AAK52094;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 639.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.

```
PR   30-NOV-2000; 2000US-0728422.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT,  Asundi V,  Zhou P,  Xu C,  Cao Y,  Ma Y;
PI   Zhao QA,  Wang D,  Wang J,  Zhang J,  Ren F,  Chen R,  Wang ZW;
PI   Xue AJ,  Yang Y,  Wejhrman T,  Goodrich R;
XX
DR   WPI; 2001-476283/51.
DR   P-PSDB; AAM78961.
XX
PT   Nucleic acids encoding polypeptides with cytokine-like activities,
PT   useful in diagnosis and gene therapy -
XX
PS   Claim 1; Page 2263; 6221pp; English.
XX
CC   The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC   encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC   (AAM80020) are omitted as the relevant pages from the sequence listing
CC   were missing at the time of publication.
XX
SQ   Sequence 601 BP; 140 A; 178 C; 147 G; 136 T; 0 other;

  Query Match             42.4%;  Score 247;  DB 22;  Length 601;
  Best Local Similarity   82.1%;  Pred. No. 5.6e-60;
  Matches  298;  Conservative    0;  Mismatches   60;  Indels    5;  Gaps    1;

Qy        221 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 280
              || |||| | || || ||| | | ||||||||| |  || ||| | || | |||||||||
Db        211 CCACTGCCTCTTCTCTTCTCGCTTGGGAACTCCAGTCTCACCTCGGCTTGCAATGGACCC 270

Qy        281 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 340
              ||||||||||||  | ||| ||||||||| ||||| ||||||||||||| |||||| |
Db        271 CAACTGCTCCTGCGAGGCTGGTGGCTCCTGCGCCTGCGCCGGCTCCTGCAAGTGCAAAAA 330

Qy        341 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCCTGTGGGCTGTGCCAA 400
              |||||||||||||||||||||||||||||||||||||||| |||||  ||||||||||||
Db        331 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGTTGCCCCCTGGGCTGTGCCAA 390

Qy        401 GTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATG 460
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        391 GTGTGCCCAGGGCTGCATCTGCAAAGGGGCGTCAGAGAAGTGCAGCTGCTGTGCCTGATG 450

Qy        461 TCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGAATTTTTTT 520
              || ||||||||||||   ||||  | | ||| ||||  | | || |  ||| |||||||
Db        451 TCGGGACAGCCCTGCTGTCAGATGAAAACAGAATGACACGTAAAATCCAGGATTTTTTTT 510

Qy        521 TCCATACAACCCTGCCCCATCTACTGTA-----TTTTTTTTAATGAAATATGTGAATGAT 575
              |   |||||| | | | ||| | ||  |     |||||||   ||||||||||||| ||
Db        511 TTTCTACAACTCCGACTCATTTGCTACATTCCTTTTTTTTCTGTGAAATATGTGAATAAT 570

Qy        576 AAT 578
              |||
Db        571 AAT 573


RESULT 9
ABN96205/c
ID   ABN96205 standard; DNA; 409 BP.
XX
AC   ABN96205;
XX
DT   13-AUG-2002  (first entry)
XX
DE   Gene #2703 used to diagnose liver cancer.
XX
KW   Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW   metastatic liver tumour; cytostatic; expression profile; disease state;
KW   disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
OS   Homo sapiens.
XX
PN   WO200229103-A2.
XX
PD   11-APR-2002.
XX
PF   02-OCT-2001; 2001WO-US30589.
XX
PR   02-OCT-2000; 2000US-237054P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Horne D,  Alvares C,  Peres-Da-Silva S,  Vockley JG;
XX
DR   WPI; 2002-426119/45.
XX
PT   Diagnosing and detecting the progression of liver cancer,
PT   hepatocellular carcinoma or metastatic liver tumor in a patient,
PT   involves detecting the level of expression of two or more genes in a
PT   liver tissue sample  -
XX
PS   Claim 1; SEQ ID NO 2703; 298pp; English.
XX
CC   The invention relates to a novel method for diagnosing and detecting the
CC   progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC   tumour in a patient, and differentiating metastatic liver cancer from
CC   hepatocellular carcinoma in a patient, involving detecting the level of
CC   expression of two or more genes represented in ABN93503-ABN97455 in a
CC   tissue sample. The method of the invention has hepatotropic, and
CC   cytostatic activity. The method is useful for diagnosing and detecting
CC   the progression of liver cancer, hepatocellular carcinoma and metastatic
CC   liver carcinoma in a patient. The method is useful for identifying
CC   expression profiles which serve as useful diagnostic markers as well as
CC   markers that can be used to monitor disease states, disease progression,
CC   drug toxicity, drug efficacy and drug metabolism.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 409 BP; 116 A; 93 C; 109 G; 87 T; 4 other;

  Query Match             42.1%;  Score 245.2;  DB 24;  Length 409;
  Best Local Similarity   88.0%;  Pred. No. 1.5e-59;
  Matches  324;  Conservative    0;  Mismatches   35;  Indels    9;  Gaps    5;

Qy        221 CCGCTGCGTGTTTTCCTCTTG-ATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACC 279
              |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db        382 CCGCTGCGTGTTTTCCTCTTGAATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACC 323

Qy        280 CCAACTGCTCCTGC-TCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAA 338
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        322 CCAACTGCTCCTGCNTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAA 263

Qy        339 GAGTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCC-TGCTGCCCTGTGGGCTGTGC 397
              |||||||||||||||||||||||||||||||||||||||| |||   ||||||||||||||
Db        262 GAGTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCNTGCNTNCCTGTGGGCTGTGC 203

Qy        398 CAAGTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTG 457
              |||||||||||||||||||||||||||||| | ||||| ||||||||||||||||||||
Db        202 CAAGTGTGCCCAGGGCTGCATCTGCAAAGGGACGTCAGACAAGTGCAGCTGCTGTGCCTG 143

Qy        458 ATGTCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGA----A 513
              ||| | |||||||  |||||  |||| || | ||||  ||||  | |||| ||||
Db        142 ATGCCAGGACAGCTGTGCTCTCAGATGTAAATAGAGCAACCTATATAAACCTGGATTTTT 83

Qy        514 TTTTTTTTCCATACAACCCTGCCCCATCTACT--GTATTTTTTTTAATGAAATATGTGAA 571
```

```
Db        82 TTTTTTTTTTGTACAACCCTGACCCGTTTGCTACATCTTTTTTTCTATGAAATATGTGAA 23

Qy       572 TGATAATA 579
Db        22 TGGCAATA 15
```

RESULT 10
ABK64607/c
ID ABK64607 standard; DNA; 409 BP.
XX
AC ABK64607;
XX
DT 18-JUN-2002 (first entry)
XX
DE Human benign prostatic hyperplasia gene #502.
XX
KW Human; benign prostatic hyperplasia; BPH; prostate cancer; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200212440-A2.
XX
PD 14-FEB-2002.
XX
PF 07-AUG-2001; 2001WO-US24708.
XX
PR 07-AUG-2000; 2000US-223323P.
PR 05-JUN-2001; 2001US-0873319.
XX
PA (GENE-) GENE LOGIC INC.
PA (NISB) JAPAN TOBACCO INC.
XX
PI Munger WE, Kulkarni P, Getzenberg RH, Waga I, Yamamoto J;
XX
DR WPI; 2002-257476/30.
XX
PT Identifying drugs for and diagnosing benign prostatic hyperplasia, by
PT detecting expression levels of one or more genes in prostate cells from
PT patient that are differentially regulated compared to normal prostate
PT cells -
XX
PS Disclosure; Page 284; 444pp; English.
XX
CC The invention relates to a method of diagnosing (I) the onset or
CC progression of benign prostatic hyperplasia (BPH), or screening (II) for
CC or identifying an agent that modulates the onset or progression of BPH.
CC The method is based on changes in gene expression in BPH tissue isolated
CC from patients exhibiting different clinical states of prostate
CC hyperplasia as compared to normal prostate tissue. (I) comprises
CC detecting the expression levels of one or more genes in prostate cells
CC from the subject that are differentially regulated compared to normal
CC prostate cells. (II) comprises preparing a first gene expression profile
CC of BPH cells or BPH-like cell population, exposing the cells to the
CC agent, preparing a second gene expression profile of the agent exposed
CC cells, and comparing the first and second gene expression profiles.
CC (I) is useful for diagnosing the onset or progression of BPH. (II) is
CC useful for identifying an agent that modulates the onset or progression
CC of BPH. The methods are useful to present information identifying
CC the expression level in a tissue or cells, by comparing the expression
CC level of genes given in the specification in the tissue or cells to the
CC level of expression of gene in the database, and displaying the
CC expression levels of at least one gene in the tissue or cell sample
CC compared to the expression level in BPH. Agents using (II) are useful for
CC treating BPH or prostate cancer. ABK64106-ABK64860 represent human
CC benign prostatic hyperplasia gene sequences of the invention.
XX
SQ Sequence 409 BP; 116 A; 93 C; 109 G; 87 T; 4 other;

Query Match 42.1%; Score 245.2; DB 24; Length 409;
Best Local Similarity 88.0%; Pred. No. 1.5e-59;
Matches 324; Conservative 0; Mismatches 35; Indels 9; Gaps 5;

```
Qy       221 CCGCTGCGTGTTTTCCTCTTG-ATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACC 279
Db       382 CCGCTGCGTGTTTTCCTCTTGAATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACC 323

Qy       280 CCAACTGCTCCTGC-TCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAA 338
Db       322 CCAACTGCTCCTGCNTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAA 263

Qy       339 GAGTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCC-TGCTGCCCTGTGGGCTGTGC 397
Db       262 GAGTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCNTGCNTNCCTGTGGGCTGTGC 203

Qy       398 CAAGTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTG 457
Db       202 CAAGTGTGCCCAGGGCTGCATCTGCAAAGGGACGTCAGACAAGTGCAGCTGCTGTGCCTG 143

Qy       458 ATGTCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGA----A 513
Db       142 ATGCCAGGACAGCTGTGCTCTCAGATGTAAATAGAGCAACCTATATAAACCTGGATTTTT 83

Qy       514 TTTTTTTTCCATACAACCCTGCCCCATCTACT--GTATTTTTTTTAATGAAATATGTGAA 571
Db        82 TTTTTTTTTTGTACAACCCTGACCCGTTTGCTACATCTTTTTTTCTATGAAATATGTGAA 23

Qy       572 TGATAATA 579
Db        22 TGGCAATA 15
```

RESULT 11
ABL62407/c
ID ABL62407 standard; DNA; 409 BP.
XX
AC ABL62407;
XX
DT 15-MAY-2002 (first entry)
XX
DE Colon adenocarcinoma related gene sequence SEQ ID NO:744.
XX
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200194629-A2.
XX
PD 13-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US10838.
XX
PR 05-JUN-2000; 2000US-209473P.
PR 05-JUN-2000; 2000US-209531P.
PR 18-SEP-2000; 2000US-233133P.
PR 18-SEP-2000; 2000US-233617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 20-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.
PR 25-SEP-2000; 2000US-234923P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-235077P.
PR 25-SEP-2000; 2000US-235082P.
PR 25-SEP-2000; 2000US-235134P.
PR 25-SEP-2000; 2000US-235280P.
PR 26-SEP-2000; 2000US-235637P.
PR 26-SEP-2000; 2000US-235638P.
PR 27-SEP-2000; 2000US-235711P.
PR 27-SEP-2000; 2000US-235720P.
PR 27-SEP-2000; 2000US-235840P.

```
PR   27-SEP-2000; 2000US-235863P.
PR   28-SEP-2000; 2000US-236028P.
PR   28-SEP-2000; 2000US-236032P.
PR   28-SEP-2000; 2000US-236033P.
PR   28-SEP-2000; 2000US-236034P.
PR   28-SEP-2000; 2000US-236109P.
PR   28-SEP-2000; 2000US-236111P.
PR   29-SEP-2000; 2000US-236842P.
PR   29-SEP-2000; 2000US-236891P.
PR   02-OCT-2000; 2000US-237172P.
PR   02-OCT-2000; 2000US-237173P.
PR   02-OCT-2000; 2000US-237278P.
PR   02-OCT-2000; 2000US-237294P.
PR   02-OCT-2000; 2000US-237295P.
PR   02-OCT-2000; 2000US-237316P.
PR   03-OCT-2000; 2000US-237425P.
PR   03-OCT-2000; 2000US-237598P.
PR   03-OCT-2000; 2000US-237604P.
PR   03-OCT-2000; 2000US-237606P.
PR   03-OCT-2000; 2000US-237608P.
PR   01-NOV-2000; 2000US-244867P.
PR   01-NOV-2000; 2000US-245084P.
XX
PA   (AVAL-) AVALON PHARM.
XX
PI   Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI   Soppet DR, Weaver Z;
XX
DR   WPI; 2002-188264/24.
XX
PT   Screening for anti-neoplastic agent involves exposing cells to a
PT   chemical agent to be tested for anti-neoplastic activity, and
PT   determining a change in expression of a gene of a signature gene set -
XX
PS   Claim 1; SEQ ID 744; 44pp; English.
XX
CC   The present invention describes a method (M1) for screening for an
CC   anti-neoplastic agent. The method involves exposing cells to a chemical
CC   agent to be tested for anti-neoplastic activity, determining a change in
CC   expression of at least one gene (I) of a signature gene set, where (I)
CC   comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC   to ABL70110), or is at least 95% identical to (S), where a change in
CC   expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC   activity and can be used in gene therapy. M1 can be used for screening
CC   an anti-neoplastic agent, and can be used for producing a product which
CC   is the data collected with respect to the anti-neoplastic agent as a
CC   result of M1, and the data is sufficient to convey the chemical
CC   structure and/or properties of the agent. M1 can be used in the
CC   treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC   oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC   adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC   infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC   carcinoma, papillary carcinoma and Wilm's tumour.
XX
SQ   Sequence 409 BP; 116 A; 93 C; 109 G; 87 T; 4 other;

  Query Match           42.1%;  Score 245.2;  DB 24;  Length 409;
  Best Local Similarity  88.0%;  Pred. No. 1.5e-59;
  Matches  324;  Conservative    0;  Mismatches   35;  Indels    9;  Gaps    5;

Qy       221 CCGCTGCGTGTTTTCCTCTTG-ATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACC 279
            ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       382 CCGCTGCGTGTTTTCCTCTTGAATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACC 323

Qy       280 CCAACTGCTCCTGC-TCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAA 338
            ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       322 CCAACTGCTCCTGCNTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAA 263

Qy       339 GAGTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCC-TGCTGCCCTGTGGGCTGTGC 397
            |||||||||||||||||||||||||||||||||||||||| |||   ||||||||||||||
Db       262 GAGTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCNTGCNTNCCTGTGGGCTGTGC 203

Qy       398 CAAGTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTG 457
            |||||||||||||||||||||||||||||| | ||||| ||||||||||||||||||||
Db       202 CAAGTGTGCCCAGGGCTGCATCTGCAAAGGGACGTCAGACAAGTGCAGCTGCTGTGCCTG 143

Qy       458 ATGTCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGA----A 513
             ||| | |||||||  |||||  |||| || | ||||  ||||  | |||| ||||
Db       142 ATGCCAGGACAGCTGTGCTCTCAGATGTAAATAGAGCAACCTATATAAACCTGGATTTTT 83

Qy       514 TTTTTTTTCCATACAACCCTGCCCCATCTACT--GTATTTTTTTTAATGAAATATGTGAA 571
             ||||||||   |||||||||| ||| | | ||   | |||||||  ||||||||||||
Db        82 TTTTTTTTTTGTACAACCCTGACCCGTTTGCTACATCTTTTTTTCTATGAAATATGTGAA 23

Qy       572 TGATAATA 579
                ||  ||||
Db        22 TGGCAATA 15


RESULT 12
ABL63983/c
ID   ABL63983 standard; DNA; 409 BP.
XX
AC   ABL63983;
XX
DT   15-MAY-2002  (first entry)
XX
DE   Breast cancer related gene sequence SEQ ID NO:2320.
XX
KW   Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW   stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW   cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200194629-A2.
XX
PD   13-DEC-2001.
XX
PF   30-MAY-2001; 2001WO-US10838.
XX
PR   05-JUN-2000; 2000US-209473P.
PR   05-JUN-2000; 2000US-209531P.
PR   18-SEP-2000; 2000US-233133P.
PR   18-SEP-2000; 2000US-233617P.
PR   20-SEP-2000; 2000US-234009P.
PR   20-SEP-2000; 2000US-234034P.
PR   20-SEP-2000; 2000US-234052P.
PR   22-SEP-2000; 2000US-234509P.
PR   22-SEP-2000; 2000US-234567P.
PR   25-SEP-2000; 2000US-234923P.
PR   25-SEP-2000; 2000US-234924P.
PR   25-SEP-2000; 2000US-235077P.
PR   25-SEP-2000; 2000US-235082P.
PR   25-SEP-2000; 2000US-235134P.
PR   25-SEP-2000; 2000US-235280P.
PR   26-SEP-2000; 2000US-235637P.
PR   26-SEP-2000; 2000US-235638P.
PR   27-SEP-2000; 2000US-235711P.
PR   27-SEP-2000; 2000US-235720P.
PR   27-SEP-2000; 2000US-235840P.
PR   27-SEP-2000; 2000US-235863P.
PR   28-SEP-2000; 2000US-236028P.
PR   28-SEP-2000; 2000US-236032P.
PR   28-SEP-2000; 2000US-236033P.
PR   28-SEP-2000; 2000US-236034P.
PR   28-SEP-2000; 2000US-236109P.
PR   28-SEP-2000; 2000US-236111P.
PR   29-SEP-2000; 2000US-236842P.
PR   29-SEP-2000; 2000US-236891P.
PR   02-OCT-2000; 2000US-237172P.
PR   02-OCT-2000; 2000US-237173P.
PR   02-OCT-2000; 2000US-237278P.
PR   02-OCT-2000; 2000US-237294P.
PR   02-OCT-2000; 2000US-237295P.
```

```
PR   02-OCT-2000; 2000US-237316P.
PR   03-OCT-2000; 2000US-237425P.
PR   03-OCT-2000; 2000US-237598P.
PR   03-OCT-2000; 2000US-237604P.
PR   03-OCT-2000; 2000US-237606P.
PR   03-OCT-2000; 2000US-237608P.
PR   01-NOV-2000; 2000US-244867P.
PR   01-NOV-2000; 2000US-245084P.
XX
PA   (AVAL-) AVALON PHARM.
XX
PI   Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI   Soppet DR, Weaver Z;
XX
DR   WPI; 2002-188264/24.
XX
PT   Screening for anti-neoplastic agent involves exposing cells to a
PT   chemical agent to be tested for anti-neoplastic activity, and
PT   determining a change in expression of a gene of a signature gene set
XX
PS   Claim 1; SEQ ID 2320; 44pp; English.
XX
CC   The present invention describes a method (M1) for screening for an
CC   anti-neoplastic agent. The method involves exposing cells to a chemical
CC   agent to be tested for anti-neoplastic activity, determining a change in
CC   expression of at least one gene (I) of a signature gene set, where (I)
CC   comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC   to ABL70110), or is at least 95% identical to (S), where a change in
CC   expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC   activity and can be used in gene therapy. M1 can be used for screening
CC   an anti-neoplastic agent, and can be used for producing a product which
CC   is the data collected with respect to the anti-neoplastic agent as a
CC   result of M1, and the data is sufficient to convey the chemical
CC   structure and/or properties of the agent. M1 can be used in the
CC   treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC   oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC   adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC   infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC   carcinoma, papillary carcinoma and Wilm's tumour.
XX
SQ   Sequence 409 BP; 116 A; 93 C; 109 G; 87 T; 4 other;

  Query Match          42.1%;  Score 245.2;  DB 24;  Length 409;
  Best Local Similarity  88.0%;  Pred. No. 1.5e-59;
  Matches  324;  Conservative    0;  Mismatches   35;  Indels    9;  Gaps    5;

Qy        221 CCGCTGCGTGTTTTCCTCTTG-ATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACC 279
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        382 CCGCTGCGTGTTTTCCTCTTGAATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACC 323

Qy        280 CCAACTGCTCCTGC-TCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAA 338
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        322 CCAACTGCTCCTGCNTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAA 263

Qy        339 GAGTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCC-TGCTGCCCTGTGGGCTGTGC 397
              |||||||||||||||||||||||||||||||||||||| ||| ||||||||||||||
Db        262 GAGTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCNTGCNTNCCTGTGGGCTGTGC 203

Qy        398 CAAGTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTG 457
              |||||||||||||||||||||||||||| | ||||| ||||||||||||||||||||
Db        202 CAAGTGTGCCCAGGGCTGCATCTGCAAAGGGACGTCAGACAAGTGCAGCTGCTGTGCCTG 143

Qy        458 ATGTCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGA----A 513
              ||| | |||||||  |||||  |||| || | ||||  ||||  | |||| ||||
Db        142 ATGCCAGGACAGCTGTGCTCTCAGATGTAAATAGAGCAACCTATATAAACCTGGATTTTT 83

Qy        514 TTTTTTTTCCATACAACCCTGCCCCATCTACT--GTATTTTTTTTAATGAAATATGTGAA 571
              ||||||||   |||||||||| ||| | | ||   | |||||||  ||||||||||||||
Db         82 TTTTTTTTTTGTACAACCCTGACCCGTTTGCTACATCTTTTTTTCTATGAAATATGTGAA 23

Qy        572 TGATAATA 579
              ||  ||||
Db         22 TGGCAATA 15
```

```
RESULT 13
ABL68231/c
ID   ABL68231 standard; DNA; 409 BP.
XX
AC   ABL68231;
XX
DT   15-MAY-2002  (first entry)
XX
DE   Kidney cancer related gene sequence SEQ ID NO:6568.
XX
KW   Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW   stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW   cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200194629-A2.
XX
PD   13-DEC-2001.
XX
PF   30-MAY-2001; 2001WO-US10838.
XX
PR   05-JUN-2000; 2000US-209473P.
PR   05-JUN-2000; 2000US-209531P.
PR   18-SEP-2000; 2000US-233133P.
PR   18-SEP-2000; 2000US-233617P.
PR   20-SEP-2000; 2000US-234009P.
PR   20-SEP-2000; 2000US-234034P.
PR   20-SEP-2000; 2000US-234052P.
PR   22-SEP-2000; 2000US-234509P.
PR   22-SEP-2000; 2000US-234567P.
PR   25-SEP-2000; 2000US-234923P.
PR   25-SEP-2000; 2000US-234924P.
PR   25-SEP-2000; 2000US-235077P.
PR   25-SEP-2000; 2000US-235082P.
PR   25-SEP-2000; 2000US-235134P.
PR   25-SEP-2000; 2000US-235280P.
PR   26-SEP-2000; 2000US-235637P.
PR   26-SEP-2000; 2000US-235638P.
PR   27-SEP-2000; 2000US-235711P.
PR   27-SEP-2000; 2000US-235720P.
PR   27-SEP-2000; 2000US-235840P.
PR   27-SEP-2000; 2000US-235863P.
PR   28-SEP-2000; 2000US-236028P.
PR   28-SEP-2000; 2000US-236032P.
PR   28-SEP-2000; 2000US-236033P.
PR   28-SEP-2000; 2000US-236034P.
PR   28-SEP-2000; 2000US-236109P.
PR   28-SEP-2000; 2000US-236111P.
PR   29-SEP-2000; 2000US-236842P.
PR   29-SEP-2000; 2000US-236891P.
PR   02-OCT-2000; 2000US-237172P.
PR   02-OCT-2000; 2000US-237173P.
PR   02-OCT-2000; 2000US-237278P.
PR   02-OCT-2000; 2000US-237294P.
PR   02-OCT-2000; 2000US-237295P.
PR   02-OCT-2000; 2000US-237316P.
PR   03-OCT-2000; 2000US-237425P.
PR   03-OCT-2000; 2000US-237598P.
PR   03-OCT-2000; 2000US-237604P.
PR   03-OCT-2000; 2000US-237606P.
PR   03-OCT-2000; 2000US-237608P.
PR   01-NOV-2000; 2000US-244867P.
PR   01-NOV-2000; 2000US-245084P.
XX
PA   (AVAL-) AVALON PHARM.
XX
PI   Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI   Soppet DR, Weaver Z;
XX
```

```
DR   WPI; 2002-188264/24.
XX
PT   Screening for anti-neoplastic agent involves exposing cells to a
PT   chemical agent to be tested for anti-neoplastic activity, and
PT   determining a change in expression of a gene of a signature gene set -
XX
PS   Claim 1; SEQ ID 6568; 44pp; English.
XX
CC   The present invention describes a method (M1) for screening for an
CC   anti-neoplastic agent. The method involves exposing cells to a chemical
CC   agent to be tested for anti-neoplastic activity, determining a change in
CC   expression of at least one gene (I) of a signature gene set, where (I)
CC   comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC   to ABL70110), or is at least 95% identical to (S), where a change in
CC   expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC   activity and can be used in gene therapy. M1 can be used for screening
CC   an anti-neoplastic agent, and can be used for producing a product which
CC   is the data collected with respect to the anti-neoplastic agent as a
CC   result of M1, and the data is sufficient to convey the chemical
CC   structure and/or properties of the agent. M1 can be used in the
CC   treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC   oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC   adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC   infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC   carcinoma, papillary carcinoma and Wilm's tumour.
XX
SQ   Sequence 409 BP; 116 A; 93 C; 109 G; 87 T; 4 other;

  Query Match             42.1%;  Score 245.2;  DB 24;  Length 409;
  Best Local Similarity   88.0%;  Pred. No. 1.5e-59;
  Matches  324;  Conservative    0;  Mismatches   35;  Indels    9;  Gaps    5;

Qy        221 CCGCTGCGTGTTTTCCTCTTG-ATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACC 279
              |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db        382 CCGCTGCGTGTTTTCCTCTTGAATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACC 323

Qy        280 CCAACTGCTCCTGC-TCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAA 338
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        322 CCAACTGCTCCTGCNTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAA 263

Qy        339 GAGTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCC-TGCTGCCCTGTGGGCTGTGC 397
              |||||||||||||||||||||||||||||||||||||| |||   |||||||||||||
Db        262 GAGTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCNTGCNTNCCTGTGGGCTGTGC 203

Qy        398 CAAGTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTG 457
              |||||||||||||||||||||||||||| | ||||| ||||||||||||||||||||
Db        202 CAAGTGTGCCCAGGGCTGCATCTGCAAAGGGACGTCAGACAAGTGCAGCTGCTGTGCCTG 143

Qy        458 ATGTCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGA----A 513
              ||| | |||||||  |||||  |||| || | ||||  ||||  | |||| ||||
Db        142 ATGCCAGGACAGCTGTGCTCTCAGATGTAAATAGAGCAACCTATATAAACCTGGATTTTT 83

Qy        514 TTTTTTTTCCATACAACCCTGCCCCATCTACT--GTATTTTTTTTAATGAAATATGTGAA 571
              ||||||||   |||||||||| ||| | | ||   | |||||||  ||||||||||||
Db         82 TTTTTTTTTTGTACAACCCTGACCCGTTTGCTACATCTTTTTTTCTATGAAATATGTGAA 23

Qy        572 TGATAATA 579
              ||  ||||
Db         22 TGGCAATA 15


RESULT 14
ABL68452/c
ID   ABL68452 standard; DNA; 409 BP.
XX
AC   ABL68452;
XX
DT   15-MAY-2002  (first entry)
XX
DE   Kidney cancer related gene sequence SEQ ID NO:6789.
XX
KW   Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW   stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW   cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200194629-A2.
XX
PD   13-DEC-2001.
XX
PF   30-MAY-2001; 2001WO-US10838.
XX
PR   05-JUN-2000; 2000US-209473P.
PR   05-JUN-2000; 2000US-209531P.
PR   18-SEP-2000; 2000US-233133P.
PR   18-SEP-2000; 2000US-233617P.
PR   20-SEP-2000; 2000US-234009P.
PR   20-SEP-2000; 2000US-234034P.
PR   20-SEP-2000; 2000US-234052P.
PR   22-SEP-2000; 2000US-234509P.
PR   22-SEP-2000; 2000US-234567P.
PR   25-SEP-2000; 2000US-234923P.
PR   25-SEP-2000; 2000US-234924P.
PR   25-SEP-2000; 2000US-235077P.
PR   25-SEP-2000; 2000US-235082P.
PR   25-SEP-2000; 2000US-235134P.
PR   25-SEP-2000; 2000US-235280P.
PR   26-SEP-2000; 2000US-235637P.
PR   26-SEP-2000; 2000US-235638P.
PR   27-SEP-2000; 2000US-235711P.
PR   27-SEP-2000; 2000US-235720P.
PR   27-SEP-2000; 2000US-235840P.
PR   27-SEP-2000; 2000US-235863P.
PR   28-SEP-2000; 2000US-236028P.
PR   28-SEP-2000; 2000US-236032P.
PR   28-SEP-2000; 2000US-236033P.
PR   28-SEP-2000; 2000US-236034P.
PR   28-SEP-2000; 2000US-236109P.
PR   28-SEP-2000; 2000US-236111P.
PR   29-SEP-2000; 2000US-236842P.
PR   29-SEP-2000; 2000US-236891P.
PR   02-OCT-2000; 2000US-237172P.
PR   02-OCT-2000; 2000US-237173P.
PR   02-OCT-2000; 2000US-237278P.
PR   02-OCT-2000; 2000US-237294P.
PR   02-OCT-2000; 2000US-237295P.
PR   02-OCT-2000; 2000US-237316P.
PR   03-OCT-2000; 2000US-237425P.
PR   03-OCT-2000; 2000US-237598P.
PR   03-OCT-2000; 2000US-237604P.
PR   03-OCT-2000; 2000US-237606P.
PR   03-OCT-2000; 2000US-237608P.
PR   01-NOV-2000; 2000US-244867P.
PR   01-NOV-2000; 2000US-245084P.
XX
PA   (AVAL-) AVALON PHARM.
XX
PI   Young PE,  Augustus M,  Carter KC,  Ebner R,  Endress G,  Horrigan S;
PI   Soppet DR,  Weaver Z;
XX
DR   WPI; 2002-188264/24.
XX
PT   Screening for anti-neoplastic agent involves exposing cells to a
PT   chemical agent to be tested for anti-neoplastic activity, and
PT   determining a change in expression of a gene of a signature gene set -
XX
PS   Claim 1; SEQ ID 6789; 44pp; English.
XX
CC   The present invention describes a method (M1) for screening for an
CC   anti-neoplastic agent. The method involves exposing cells to a chemical
CC   agent to be tested for anti-neoplastic activity, determining a change in
CC   expression of at least one gene (I) of a signature gene set, where (I)
CC   comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC   to ABL70110), or is at least 95% identical to (S), where a change in
```

```
CC   expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC   activity and can be used in gene therapy. M1 can be used for screening
CC   an anti-neoplastic agent, and can be used for producing a product which
CC   is the data collected with respect to the anti-neoplastic agent as a
CC   result of M1, and the data is sufficient to convey the chemical
CC   structure and/or properties of the agent. M1 can be used in the
CC   treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC   oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC   adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC   infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC   carcinoma, papillary carcinoma and Wilm's tumour.
XX
SQ   Sequence 409 BP; 116 A; 93 C; 109 G; 87 T; 4 other;

  Query Match          42.1%;  Score 245.2;  DB 24;  Length 409;
  Best Local Similarity   88.0%;  Pred. No. 1.5e-59;
  Matches  324;  Conservative    0;  Mismatches   35;  Indels    9;  Gaps    5;

Qy        221 CCGCTGCGTGTTTTCCTCTTG-ATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACC 279
              ||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db        382 CCGCTGCGTGTTTTCCTCTTGAATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACC 323

Qy        280 CCAACTGCTCCTGC-TCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAA 338
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        322 CCAACTGCTCCTGCNTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAA 263

Qy        339 GAGTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCC-TGCTGCCCTGTGGGCTGTGC 397
              ||||||||||||||||||||||||||||||||||||||| ||| ||||||||||||||
Db        262 GAGTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCNTGCNTNCCTGTGGGCTGTGC 203

Qy        398 CAAGTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTG 457
              |||||||||||||||||||||||||||||| | ||||| ||||||||||||||||||||
Db        202 CAAGTGTGCCCAGGGCTGCATCTGCAAAGGGACGTCAGACAAGTGCAGCTGCTGTGCCTG 143

Qy        458 ATGTCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGA----A 513
              ||| | |||||||  |||||  |||| || | ||||  ||||  | |||| ||||
Db        142 ATGCCAGGACAGCTGTGCTCTCAGATGTAAATAGAGCAACCTATATAAACCTGGATTTTT 83

Qy        514 TTTTTTTTCCATACAACCCTGCCCCATCTACT--GTATTTTTTTTAATGAAATATGTGAA 571
              ||||||||  |||||||||| ||| | | ||   | |||||||  ||||||||||||||
Db         82 TTTTTTTTTTGTACAACCCTGACCCGTTTGCTACATCTTTTTTTCTATGAAATATGTGAA 23

Qy        572 TGATAATA 579
              ||  ||||
Db         22 TGGCAATA 15


RESULT 15
ABL68852/c
ID   ABL68852 standard; DNA; 409 BP.
XX
AC   ABL68852;
XX
DT   15-MAY-2002  (first entry)
XX
DE   Kidney cancer related gene sequence SEQ ID NO:7189.
XX
KW   Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW   stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW   cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200194629-A2.
XX
PD   13-DEC-2001.
XX
PF   30-MAY-2001; 2001WO-US10838.
XX
PR   05-JUN-2000; 2000US-209473P.
PR   05-JUN-2000; 2000US-209531P.
PR   18-SEP-2000; 2000US-233133P.
PR   18-SEP-2000; 2000US-233617P.
PR   20-SEP-2000; 2000US-234009P.
PR   20-SEP-2000; 2000US-234034P.
PR   20-SEP-2000; 2000US-234052P.
PR   22-SEP-2000; 2000US-234509P.
PR   22-SEP-2000; 2000US-234567P.
PR   25-SEP-2000; 2000US-234923P.
PR   25-SEP-2000; 2000US-234924P.
PR   25-SEP-2000; 2000US-235077P.
PR   25-SEP-2000; 2000US-235082P.
PR   25-SEP-2000; 2000US-235134P.
PR   25-SEP-2000; 2000US-235280P.
PR   26-SEP-2000; 2000US-235637P.
PR   26-SEP-2000; 2000US-235638P.
PR   27-SEP-2000; 2000US-235711P.
PR   27-SEP-2000; 2000US-235720P.
PR   27-SEP-2000; 2000US-235840P.
PR   27-SEP-2000; 2000US-235863P.
PR   28-SEP-2000; 2000US-236028P.
PR   28-SEP-2000; 2000US-236032P.
PR   28-SEP-2000; 2000US-236033P.
PR   28-SEP-2000; 2000US-236034P.
PR   28-SEP-2000; 2000US-236109P.
PR   28-SEP-2000; 2000US-236111P.
PR   29-SEP-2000; 2000US-236842P.
PR   29-SEP-2000; 2000US-236891P.
PR   02-OCT-2000; 2000US-237172P.
PR   02-OCT-2000; 2000US-237173P.
PR   02-OCT-2000; 2000US-237278P.
PR   02-OCT-2000; 2000US-237294P.
PR   02-OCT-2000; 2000US-237295P.
PR   02-OCT-2000; 2000US-237316P.
PR   03-OCT-2000; 2000US-237425P.
PR   03-OCT-2000; 2000US-237598P.
PR   03-OCT-2000; 2000US-237604P.
PR   03-OCT-2000; 2000US-237606P.
PR   03-OCT-2000; 2000US-237608P.
PR   01-NOV-2000; 2000US-244867P.
PR   01-NOV-2000; 2000US-245084P.
XX
PA   (AVAL-) AVALON PHARM.
XX
PI   Young PE,  Augustus M,  Carter KC,  Ebner R,  Endress G,  Horrigan S;
PI   Soppet DR,  Weaver Z;
XX
DR   WPI; 2002-188264/24.
XX
PT   Screening for anti-neoplastic agent involves exposing cells to a
PT   chemical agent to be tested for anti-neoplastic activity, and
PT   determining a change in expression of a gene of a signature gene set
XX
PS   Claim 1; SEQ ID 7189; 44pp; English.
XX
CC   The present invention describes a method (M1) for screening for an
CC   anti-neoplastic agent. The method involves exposing cells to a chemical
CC   agent to be tested for anti-neoplastic activity, determining a change in
CC   expression of at least one gene (I) of a signature gene set, where (I)
CC   comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC   to ABL70110), or is at least 95% identical to (S), where a change in
CC   expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC   activity and can be used in gene therapy. M1 can be used for screening
CC   an anti-neoplastic agent, and can be used for producing a product which
CC   is the data collected with respect to the anti-neoplastic agent as a
CC   result of M1, and the data is sufficient to convey the chemical
CC   structure and/or properties of the agent. M1 can be used in the
CC   treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC   oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC   adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC   infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC   carcinoma, papillary carcinoma and Wilm's tumour.
XX
SQ   Sequence 409 BP; 116 A; 93 C; 109 G; 87 T; 4 other;
```

Query Match 42.1%; Score 245.2; DB 24; Length 409;
Best Local Similarity 88.0%; Pred. No. 1.5e-59;
Matches 324; Conservative 0; Mismatches 35; Indels 9; Gaps 5;

```
Qy        221 CCGCTGCGTGTTTTCCTCTTG-ATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACC 279
              ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        382 CCGCTGCGTGTTTTCCTCTTGAATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACC 323

Qy        280 CCAACTGCTCCTGC-TCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAA 338
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        322 CCAACTGCTCCTGCNTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAA 263

Qy        339 GAGTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCC-TGCTGCCCTGTGGGCTGTGC 397
              |||||||||||||||||||||||||||||||||||||| ||| ||||||||||||||||
Db        262 GAGTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCNTGCNTNCCTGTGGGCTGTGC 203

Qy        398 CAAGTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTG 457
              |||||||||||||||||||||||||||||| | ||||| ||||||||||||||||||||
Db        202 CAAGTGTGCCCAGGGCTGCATCTGCAAAGGGACGTCAGACAAGTGCAGCTGCTGTGCCTG 143

Qy        458 ATGTCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGA----A 513
              ||| | |||||||  |||||  |||| || | ||||  ||||  | |||| ||||
Db        142 ATGCCAGGACAGCTGTGCTCTCAGATGTAAATAGAGCAACCTATATAAACCTGGATTTTT 83

Qy        514 TTTTTTTTCCATACAACCCTGCCCCATCTACT--GTATTTTTTTTAATGAAATATGTGAA 571
              ||||||||   |||||||||| ||| | | ||   | |||||||  ||||||||||||||
Db         82 TTTTTTTTTTGTACAACCCTGACCCGTTTGCTACATCTTTTTTTCTATGAAATATGTGAA 23

Qy        572 TGATAATA 579
              ||  ||||
Db         22 TGGCAATA 15
```

Search completed: July 16, 2003, 11:09:59
Job time : 101.587 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 16, 2003, 08:51:42 ; Search time 636.456 Seconds
(without alignments)
14835.236 Million cell updates/sec

Title: US-09-981-353-114
Perfect score: 583
Sequence: 1 ccctgtaggtgctgtccttg.....tatgtgaatgataatagtca 583

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	304.4	52.2	376	14	BM694421	BM694421 UI-E-CI1-
2	292.4	50.2	516	12	BG566970	BG566970 602589480
3	286.8	49.2	406	10	AW952177	AW952177 EST364247
4	282.8	48.5	399	14	BQ440401	BQ440401 AGENCOURT
5	280.8	48.2	499	13	BM458221	BM458221 AGENCOURT
6	279.8	48.0	410	14	BQ428187	BQ428187 AGENCOURT
7	278.2	47.7	1509	14	BM924728	BM924728 AGENCOURT
8	276.8	47.5	513	12	BG900215	BG900215 HOA36-1-B
9	276.8	47.5	568	13	BG926575	BG926575 HNC56-1-G
10	276.8	47.5	575	13	BG926326	BG926326 HNC55-1-G
11	276.8	47.5	1074	14	BM921692	BM921692 AGENCOURT
12	276.8	47.5	1159	13	BM544354	BM544354 AGENCOURT
13	275.8	47.3	493	12	BG897617	BG897617 HOA15-1-B
14	275.2	47.2	576	13	BG925124	BG925124 HNC37-1-B
15	274.8	47.1	407	10	AV661102	AV661102 AV661102
16	274.8	47.1	439	10	AV660814	AV660814 AV660814
17	274.8	47.1	639	13	BG925334	BG925334 HNC38-1-H
18	274.6	47.1	352	14	F32133	F32133 HSPD24679 H
19	273.6	46.9	467	12	BF526059	BF526059 602071168
20	273.2	46.9	618	13	BG928106	BG928106 HNC18-1-A
21	272.4	46.7	445	12	BG899081	BG899081 HOA21-1-F
22	272.2	46.7	518	13	BG927843	BG927843 HNC66-1-E
23	271.8	46.6	705	12	BF031192	BF031192 601558996
24	271.4	46.6	828	12	BG536644	BG536644 602566265
25	270.8	46.4	414	14	F26428	F26428 HSPD13906 H
26	270.6	46.4	323	14	F26407	F26407 HSPD13872 H
27	270.6	46.4	454	12	BG564723	BG564723 602589940
28	270.6	46.4	563	12	BF676799	BF676799 602084361
29	270.6	46.4	722	12	BG566693	BG566693 602585752
30	270.6	46.4	744	12	BG613385	BG613385 602641273
31	270.6	46.4	917	12	BF130769	BF130769 601818416
32	270	46.3	734	13	BI759852	BI759852 603045848
33	270	46.3	935	12	BG616113	BG616113 602643577
34	269.6	46.2	426	12	BG900998	BG900998 HOA52-1-E
35	269	46.1	958	12	BG533343	BG533343 601860726
36	268.2	46.0	503	10	AV662264	AV662264 AV662264
37	267.2	45.8	1186	12	BG568840	BG568840 602588006
38	266.4	45.7	976	12	BG540271	BG540271 602569164
39	265.8	45.6	332	14	BM847755	BM847755 K-EST0127
40	265.6	45.6	301	14	F27787	F27787 HSPD16007 H
41	264.4	45.4	442	13	BG927047	BG927047 HNC25-1-G
42	263.2	45.1	317	14	F26137	F26137 HSPD13466 H
43	262.6	45.0	315	14	F24837	F24837 HSPD11483 H
44	261.2	44.8	296	14	F30476	F30476 HSPD20959 H
45	261.2	44.8	461	12	BF766332	BF766332 IL2-CS004

ALIGNMENTS

RESULT 1
BM694421
LOCUS BM694421 376 bp mRNA linear EST 28-FEB-2002
DEFINITION UI-E-CI1-afq-d-21-0-UI.r1 UI-E-CI1 Homo sapiens cDNA clone UI-E-CI1-afq-d-21-0-UI 5', mRNA sequence.
ACCESSION BM694421
VERSION BM694421.1 GI:19007679
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 376)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, Univeristy of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, Univeristy of Iowa
DNA Sequencing by: Dr. M. Bento Soares, Univeristy of Iowa

Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).
Seq primer: M13 Reverse.

FEATURES Location/Qualifiers
source 1. .376
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-E-CI1-afq-d-21-0-UI"
/clone_lib="UI-E-CI1"
/tissue_type="RPE and Choroid"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; UI-E-CI1 is a normalized cDNA library containing the following tissue(s): RPE and Choroid. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is ACCTA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

BASE COUNT 92 a 104 c 85 g 95 t
ORIGIN

Query Match 52.2%; Score 304.4; DB 14; Length 376;
Best Local Similarity 91.2%; Pred. No. 1.7e-72;
Matches 323; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

```
Qy        226 GCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCCCAACT 285
              || | |   | ||||| | |||| |||| | || ||  | ||||||||||||||||||||
Db         10 GCCTCTCAACTTCTTGCTTGGGATCTCCAACCTCACCGCGGCTCGAAATGGACCCCAACT 69

Qy        286 GCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGAGTGCA 345
              |||||||| |  ||| |||||||||  ||||  | ||||||||||||||||||||||||
Db         70 GCTCCTGCGCCACTGGTGGCTCCTGCACCTGCACTGGCTCCTGCAAATGCAAAGAGTGCA 129

Qy        346 AATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAAGTGTG 405
              |||||| ||||||||||||||||||||||||||||||||  || ||||||||||||||||
Db        130 AATGCAACTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCCATGAGCTGTGCCAAGTGTG 189

Qy        406 CCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATGTCCGG 465
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        190 CCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATGTCCGG 249

Qy        466 ACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGAATTTTTTTTCCAT 525
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        250 ACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGAATTTTTTTTCCAT 309

Qy        526 ACAACCCTGCCCCATCTACTGTATTTTTTTTAATGAAATATGTGAATGATAATA 579
              ||||||||| ||||| ||||||||||||||||||||||||||||||||||||
Db        310 ACAACCCTGACCCATTTACTGTATTTTTTTTAATGAAATATGTGAATGATAATA 363
```

RESULT 2
BG566970
LOCUS BG566970 516 bp mRNA linear EST 10-APR-2001
DEFINITION 602589480F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4723818 5', mRNA sequence.
ACCESSION BG566970
VERSION BG566970.1 GI:13574623
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 516)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1583 row: o column: 19
High quality sequence stop: 483.

FEATURES Location/Qualifiers
source 1. .516
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4723818"
/clone_lib="NIH_MGC_76"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.85 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."

BASE COUNT 174 a 122 c 110 g 110 t
ORIGIN

Query Match 50.2%; Score 292.4; DB 12; Length 516;
Best Local Similarity 91.0%; Pred. No. 3.6e-69;
Matches 322; Conservative 0; Mismatches 31; Indels 1; Gaps 1;

```
Qy        226 GCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCCCAACT 285
              || | |   | ||||| | |||| |||| | || ||  | ||||||||||||||||||||
Db         25 GCCTCTCAACTTCTTGCTTGGGATCTCCAACCTCACCGCGGCTCGAAATGGACCCCAACT 84

Qy        286 GCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGAGTGCA 345
              |||||||| |  ||| |||||||||  ||||  | ||||||||||||||||||||||||
Db         85 GCTCCTGCGCCACTGGTGGCTCCTGCACCTGCACTGGCTCCTGCAAATGCAAAGAGTGCA 144

Qy        346 AATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAAGTGTG 405
              |||||| ||||||||||||||||||||||||||||||||  || ||||||||||||||||
Db        145 AATGCAACTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCCATGAGCTGTGCCAAGTGTG 204

Qy        406 CCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATGTCCGG 465
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        205 CCCA-GGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATGTCCGG 263

Qy        466 ACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGAATTTTTTTTCCAT 525
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        264 ACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGAATTTTTTTTCCAT 323

Qy        526 ACAACCCTGCCCCATCTACTGTATTTTTTTTAATGAAATATGTGAATGATAATA 579
              ||||||||| ||||| ||||||||||||||||||||||||||||||||||||
Db        324 ACAACCCTGACCCATTTACTGTATTTTTTTTAATGAAATATGTGAATGATAATA 377
```

RESULT 3
AW952177
LOCUS AW952177 406 bp mRNA linear EST 01-JUN-2000
DEFINITION EST364247 MAGE resequences, MAGB Homo sapiens cDNA, mRNA sequence.
ACCESSION AW952177
VERSION AW952177.1 GI:8141859
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 406)
  AUTHORS   Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
            ,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
            Quackenbush,J.
  TITLE     Assessment of gene expression patterns in a model of colon tumor
            metastasis using a 19,200 element cDNA microarray
  JOURNAL   Unpublished (2000)
COMMENT     Contact: John Quackenbush
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 3528
            Fax: 301 838 0208
            Email: johnq@tigr.org
            Plate: 36
            Seq primer: Reverse.
FEATURES             Location/Qualifiers
     source          1. .406
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="MAGE resequences, MAGB"
                     /note="Vector: pBluescriptSKm"
BASE COUNT      103 a    114 c     96 g     93 t
ORIGIN

  Query Match             49.2%;  Score 286.8;  DB 10;  Length 406;
  Best Local Similarity   88.1%;  Pred. No. 1.1e-67;
  Matches  312;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy        226 GCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCCCAACT 285
              || | |   | ||||| | |||| || |  | || ||  | ||||||||||||||||||
Db          9 GCCTCTCAACTTCTTGCTTGGGATCTTCAACCTCACCGCGGCTCGAAATGGACCCCAACT 68

Qy        286 GCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGAGTGCA 345
              |||||||| |  ||| |||||||||  |||| | ||||||||||||||||| |||||||
Db         69 GCTCCTGCGCCACTGGTGGCTCCTGCACCTGCACTGGCTCCTGCAAATGCAACGAGTGCA 128

Qy        346 AATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCCTGTGGGCTGTGCCAAGTGTG 405
              |||||||||||||||||||||||||||||||||||||||||||  || |||| ||||||||||
Db        129 AATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCCATGAGCTGAGCCAAGTGTG 188

Qy        406 CCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATGTCCGG 465
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        189 CCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATGTCCGG 248

Qy        466 ACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGAATTTTTTTTCCAT 525
              |||||||||||||||||||| || |||||||||||||||||||||| |||||| |||
Db        249 ACAGCCCTGCTCGAAGATATATAACGAGTGACCTGCACAAACTTGGAATCTTTTTTGCAT 308

Qy        526 ACAACCCTGCCCCATCTACTGTATTTTTTTTAATGAAATATGTGAATGATAATA 579
              |||||||| |||||  |||| ||   |||||||||||||||||||||||||
Db        309 ACAACCCTGACCCATTAACTGAATCCCTTTTAATGAAATATGTGAATGATAATA 362


RESULT 4
BQ440401
LOCUS       BQ440401                399 bp    mRNA    linear   EST 24-MAY-2002
DEFINITION  AGENCOURT_7896045 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:6155996
            5', mRNA sequence.
ACCESSION   BQ440401
VERSION     BQ440401.1  GI:21179477
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 399)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13499  row: i  column: 21
            High quality sequence stop: 398.
FEATURES             Location/Qualifiers
     source          1. .399
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6155996"
                     /clone_lib="NIH_MGC_71"
                     /tissue_type="leiomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 2.1 kb.  "
BASE COUNT       97 a    104 c     88 g    110 t
ORIGIN

  Query Match             48.5%;  Score 282.8;  DB 14;  Length 399;
  Best Local Similarity   89.8%;  Pred. No. 1.4e-66;
  Matches  327;  Conservative    0;  Mismatches   32;  Indels    5;  Gaps    2;

Qy        221 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 61

Qy        281 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 121

Qy        341 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCCTGTGGGCTGTGCCAA 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCCTGTGGGCTGTGCCAA 181

Qy        401 GTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATG 460
              |||||||||||||||||||||||||||| | ||||| ||||||||||||||||||||||
Db        182 GTGTGCCCAGGGCTGCATCTGCAAAGGGACGTCAGACAAGTGCAGCTGCTGTGCCTGATG 241

Qy        461 TCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGA---ATTTT 517
              | ||||||| |||||  |||| || | ||||  ||||  | |||| ||||    ||||
Db        242 CCAGGACAGCTGTGCTCTCAGATGTAAATAGAGCAACCTATATAAACCTGGATTTTTTTT 301

Qy        518 TTTTCCATACAACCCTGCCCCATCTACT--GTATTTTTTTTAATGAAATATGTGAATGAT 575
              ||||  |||||||||| ||| | | ||   | |||||||  ||||||||||||||||
Db        302 TTTTTTGTACAACCCTGACCCGTTTGCTACATCTTTTTTTCTATGAAATATGTGAATGGC 361

Qy        576 AATA 579
              ||||
Db        362 AATA 365


RESULT 5
BM458221
LOCUS       BM458221                499 bp    mRNA    linear   EST 05-FEB-2002
DEFINITION  AGENCOURT_6411456 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5530594
            5', mRNA sequence.
ACCESSION   BM458221
VERSION     BM458221.1  GI:18507261
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 499)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
```

```
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12210  row: o  column: 11
            High quality sequence start: 76
            High quality sequence stop: 498.
FEATURES             Location/Qualifiers
     source          1. .499
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5530594"
                     /clone_lib="NIH_MGC_71"
                     /tissue_type="leiomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 2.1 kb.  "
BASE COUNT      121 a    136 c    122 g    120 t
ORIGIN

  Query Match             48.2%;  Score 280.8;  DB 13;  Length 499;
  Best Local Similarity   89.3%;  Pred. No. 5.3e-66;
  Matches  327;  Conservative    0;  Mismatches   32;  Indels    7;  Gaps    2;

Qy        221 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         98 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 157

Qy        281 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        158 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 217

Qy        341 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAA 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        218 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAA 277

Qy        401 GTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATG 460
              |||||||||||||||||||||||||| | ||||| ||||||||||||||||||||||||
Db        278 GTGTGCCCAGGGCTGCATCTGCAAAGGGACGTCAGACAAGTGCAGCTGCTGTGCCTGATG 337

Qy        461 TCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGA-----ATT 515
              | ||||||| |||||  |||| || | ||||  ||||  | |||| ||||      ||
Db        338 CCAGGACAGCTGTGCTCTCAGATGTAAATAGAGCAACCTATATAAACCTGGATTTTTTTT 397

Qy        516 TTTTTTCCATACAACCCTGCCCCATCTACT--GTATTTTTTTTAATGAAATATGTGAATG 573
              ||||||   |||||||||| ||| | | ||   | |||||||  ||||||||||||||||
Db        398 TTTTTTTTGTACAACCCTGACCCGTTTGCTACATCTTTTTTTCTATGAAATATGTGAATG 457

Qy        574 ATAATA 579
                ||||
Db        458 GCAATA 463


RESULT 6
BQ428187
LOCUS       BQ428187                 410 bp    mRNA    linear   EST 24-MAY-2002
DEFINITION  AGENCOURT_7890364 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:6176804
            5', mRNA sequence.
ACCESSION   BQ428187
VERSION     BQ428187.1  GI:21167263
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 410)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13553  row: l  column: 21
            High quality sequence stop: 409.
FEATURES             Location/Qualifiers
     source          1. .410
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6176804"
                     /clone_lib="NIH_MGC_71"
                     /tissue_type="leiomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 2.1 kb.  "
BASE COUNT      102 a    104 c     89 g    115 t
ORIGIN

  Query Match             48.0%;  Score 279.8;  DB 14;  Length 410;
  Best Local Similarity   89.1%;  Pred. No. 9.4e-66;
  Matches  327;  Conservative    0;  Mismatches   32;  Indels    8;  Gaps    2;

Qy        221 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          5 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 64

Qy        281 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         65 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 124

Qy        341 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAA 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        125 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAA 184

Qy        401 GTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATG 460
              |||||||||||||||||||||||||| | ||||| ||||||||||||||||||||||||
Db        185 GTGTGCCCAGGGCTGCATCTGCAAAGGGACGTCAGACAAGTGCAGCTGCTGTGCCTGATG 244

Qy        461 TCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGA------AT 514
              | ||||||| |||||  |||| || | ||||  ||||  | |||| ||||      |
Db        245 CCAGGACAGCTGTGCTCTCAGATGTAAATAGAGCAACCTATATAAACCTGGATTTTTTTT 304

Qy        515 TTTTTTTCCATACAACCCTGCCCCATCTACT--GTATTTTTTTTAATGAAATATGTGAAT 572
              |||||||   |||||||||| ||| | | ||   | |||||||  ||||||||||||||
Db        305 TTTTTTTTTGTACAACCCTGACCCGTTTGCTACATCTTTTTTTCTATGAAATATGTGAAT 364

Qy        573 GATAATA 579
                | ||||
Db        365 GGCAATA 371


RESULT 7
BM924728
LOCUS       BM924728                 1509 bp    mRNA    linear   EST 12-MAR-2002
DEFINITION  AGENCOURT_6653773 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5761162
            5', mRNA sequence.
ACCESSION   BM924728
VERSION     BM924728.1  GI:19375107
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1509)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
```

```
  JOURNAL    Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Life Technologies, Inc.
              cDNA Library Preparation: Life Technologies, Inc.
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Agencourt Bioscience Corporation
              Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM12809  row: f  column: 11
             High quality sequence stop: 486.
FEATURES               Location/Qualifiers
     source            1. .1509
                       /organism="Homo sapiens"
                       /db_xref="taxon:9606"
                       /clone="IMAGE:5761162"
                       /clone_lib="NIH_MGC_116"
                       /lab_host="DH10B"
                       /note="Organ: pooled colon, kidney, stomach; Vector:
                       pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
                       source anonymous pool of 3 colons, age 26 yo male, 49 yo
                       female, 71 yo male colon; 46 yo male kidney, and pool of 2
                       stomachs, 62 yo male and 70 yo female.  Library is
                       oligo-dT primed and  directionally cloned (EcoRV site is
                       destroyed upon cloning).  Average insert size 1.4 kb,
                       insert size range 1-3 kb.  Library is normalized and
                       enriched for full-length clones and was constructed by C.
                       Gruber  (Invitrogen).  Research Genetics tracking code
                       023. Note: this is a NIH_MGC Library."
BASE COUNT      494 a    450 c    264 g    298 t      3 others
ORIGIN

  Query Match             47.7%;  Score 278.2;  DB 14;  Length 1509;
  Best Local Similarity   87.3%;  Pred. No. 3.8e-65;
  Matches  331;  Conservative    0;  Mismatches   38;  Indels   10;  Gaps    2;

Qy        211 CCAGCTCGAGCCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCG 270
              || |   | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 CCGGGATGTCCCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCG 61

Qy        271 AAATGGACCCCAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCA 330
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 AAATGGACCCCAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCA 121

Qy        331 AATGCAAAGAGTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGG 390
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 AATGCAAAGAGTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGG 181

Qy        391 GCTGTGCCAAGTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCT 450
              |||||||||||||||||||||||||||||||||| | ||||| ||||||||||||||
Db        182 GCTGTGCCAAGTGTGCCCAGGGCTGCATCTGCAAAGGGACGTCAGACAAGTGCAGCTGCT 241

Qy        451 GTGCCTGATGTCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTG 510
              |||||||||| | ||||||| |||||  |||| || | ||||  ||||  | |||| ||
Db        242 GTGCCTGATGCCAGGACAGCTGTGCTCTCAGATGTAAATAGAGCAACCTATATAAACCTG 301

Qy        511 GA--------ATTTTTTTTCCATACAACCCTGCCCCATCTACT--GTATTTTTTTTAATG 560
              ||         ||||||||   |||||||||| ||| | | ||   | |||||||  |||
Db        302 GATTTTTTTTTTTTTTTTTTTGTACAACCCTGACCCGTTTGCTACATCTTTTTTTCTATG 361

Qy        561 AAATATGTGAATGATAATA 579
              |||||||||||||  ||||
Db        362 AAATATGTGAATGGCAATA 380


RESULT 8
BG900215
LOCUS       BG900215                 513 bp    mRNA    linear   EST 06-NOV-2001
DEFINITION  HOA36-1-B9 HOA (Human Osteoarthritic Cartilage) Homo sapiens cDNA,
            mRNA sequence.
ACCESSION   BG900215
VERSION     BG900215.1  GI:14310464
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 513)
  AUTHORS   Kumar,S., Connor,J.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J.,
            Sathe,G., Mui,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and
            Lark,M.W.
  TITLE     Identification and initial characterization of 5000 expressed
            sequenced tags (ESTs) each from adult human normal and
            osteoarthritic cartilage cDNA libraries
  JOURNAL   Osteoarthr. Cartil. 9 (7), 641-653 (2001)
  MEDLINE   21482651
COMMENT     Contact: Sanjay Kumar
            UW2109
            GlaxoSmithKline
            709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA
            Tel: 610-270-7245
            Fax: 610-270-5598
            Email: sanjay_kumar-1@gsk.com
            Seq primer: T7.
FEATURES               Location/Qualifiers
     source            1. .513
                       /organism="Homo sapiens"
                       /db_xref="taxon:9606"
                       /clone_lib="HOA (Human Osteoarthritic Cartilage)"
                       /tissue_type="cartilage"
                       /lab_host="E.coli DH10 B"
                       /note="Vector: pSPORT I; Site_1: SalI; Site_2: NotI;
                       Directional"
BASE COUNT      141 a    130 c    110 g    132 t
ORIGIN

  Query Match             47.5%;  Score 276.8;  DB 12;  Length 513;
  Best Local Similarity   88.4%;  Pred. No. 6.6e-65;
  Matches  327;  Conservative    0;  Mismatches   32;  Indels   11;  Gaps    2;

Qy        221 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         85 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 144

Qy        281 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        145 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 204

Qy        341 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAA 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        205 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAA 264

Qy        401 GTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATG 460
              |||||||||||||||||||||||||||| | ||||| |||||||||||||||||||||||
Db        265 GTGTGCCCAGGGCTGCATCTGCAAAGGGACGTCAGACAAGTGCAGCTGCTGTGCCTGATG 324

Qy        461 TCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGA-------- 512
              | |||||||  |||||  |||| || | ||||  ||||  | |||| ||||
Db        325 CCAGGACAGCTGTGCTCTCAGATGTAAATAGAGCAACCTATATAAACCTGGATTTTTTTT 384

Qy        513 -ATTTTTTTTCCATACAACCCTGCCCCATCTACT--GTATTTTTTTTAATGAAATATGTG 569
               ||||||||   |||||||||| ||| | | ||   | |||||||  ||||||||||||
Db        385 TTTTTTTTTTTTGTACAACCCTGACCCGTTTGCTACATCTTTTTTTCTATGAAATATGTG 444

Qy        570 AATGATAATA 579
              ||||  ||||
Db        445 AATGGCAATA 454


RESULT 9
BG926575
LOCUS       BG926575                 568 bp    mRNA    linear   EST 06-NOV-2001
DEFINITION  HNC56-1-G9.R HNC (Human Normal Cartilage) Homo sapiens cDNA, mRNA
            sequence.
ACCESSION   BG926575
```

```
VERSION     BG926575.1  GI:14321098
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 568)
  AUTHORS   Kumar,S., Connor,J.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J.,
            Sathe,G., Mui,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and
            Lark,M.W.
  TITLE     Identification and initial characterization of 5000 expressed
            sequenced tags (ESTs) each from adult human normal and
            osteoarthritic cartilage cDNA libraries
  JOURNAL   Osteoarthr. Cartil. 9 (7), 641-653 (2001)
  MEDLINE   21482651
COMMENT     Contact: Sanjay Kumar
            UW2109
            GlaxoSmithKline
            709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA
            Tel: 610-270-7245
            Fax: 610-270-5598
            Email: sanjay_kumar-1@gsk.com
            Seq primer: T7.
FEATURES             Location/Qualifiers
     source          1. .568
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="HNC (Human Normal Cartilage)"
                     /tissue_type="cartilage"
                     /lab_host="E.coli DH10 B"
                     /note="Vector: pSPORT I; Site_1: SalI; Site_2: NotI;
                     Directional"
BASE COUNT      146 a    147 c    126 g    149 t
ORIGIN

  Query Match             47.5%;  Score 276.8;  DB 13;  Length 568;
  Best Local Similarity   88.4%;  Pred. No. 6.8e-65;
  Matches  327;  Conservative    0;  Mismatches   32;  Indels   11;  Gaps    2;

Qy        221 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         84 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 143

Qy        281 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        144 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 203

Qy        341 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAA 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        204 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAA 263

Qy        401 GTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATG 460
              |||||||||||||||||||||||||| | ||||| ||||||||||||||||||||||||
Db        264 GTGTGCCCAGGGCTGCATCTGCAAAGGGACGTCAGACAAGTGCAGCTGCTGTGCCTGATG 323

Qy        461 TCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGA-------- 512
               | ||||||| |||||  |||| || | ||||  |||| | |||| ||||
Db        324 CCAGGACAGCTGTGCTCTCAGATGTAAATAGAGCAACCTATATAAACCTGGATTTTTTTT 383

Qy        513 -ATTTTTTTTCCATACAACCCTGCCCCATCTACT--GTATTTTTTTTAATGAAATATGTG 569
               ||||||||   |||||||||| ||| | | ||   | |||||||  ||||||||||||
Db        384 TTTTTTTTTTTTGTACAACCCTGACCCGTTTGCTACATCTTTTTTTCTATGAAATATGTG 443

Qy        570 AATGATAATA 579
              ||||  ||||
Db        444 AATGGCAATA 453


RESULT 10
BG926326
LOCUS       BG926326                575 bp    mRNA    linear   EST 06-NOV-2001
DEFINITION  HNC55-1-G8.R HNC (Human Normal Cartilage) Homo sapiens cDNA, mRNA
            sequence.
```

```
ACCESSION   BG926326
VERSION     BG926326.1  GI:14320849
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 575)
  AUTHORS   Kumar,S., Connor,J.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J.,
            Sathe,G., Mui,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and
            Lark,M.W.
  TITLE     Identification and initial characterization of 5000 expressed
            sequenced tags (ESTs) each from adult human normal and
            osteoarthritic cartilage cDNA libraries
  JOURNAL   Osteoarthr. Cartil. 9 (7), 641-653 (2001)
  MEDLINE   21482651
COMMENT     Contact: Sanjay Kumar
            UW2109
            GlaxoSmithKline
            709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA
            Tel: 610-270-7245
            Fax: 610-270-5598
            Email: sanjay_kumar-1@gsk.com
            Seq primer: T7.
FEATURES             Location/Qualifiers
     source          1. .575
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="HNC (Human Normal Cartilage)"
                     /tissue_type="cartilage"
                     /lab_host="E.coli DH10 B"
                     /note="Vector: pSPORT I; Site_1: SalI; Site_2: NotI;
                     Directional"
BASE COUNT      147 a    153 c    128 g    147 t
ORIGIN

  Query Match             47.5%;  Score 276.8;  DB 13;  Length 575;
  Best Local Similarity   88.4%;  Pred. No. 6.9e-65;
  Matches  327;  Conservative    0;  Mismatches   32;  Indels   11;  Gaps    2;

Qy        221 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         91 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 150

Qy        281 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        151 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 210

Qy        341 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAA 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        211 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAA 270

Qy        401 GTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATG 460
              |||||||||||||||||||||||||| | ||||| ||||||||||||||||||||||||
Db        271 GTGTGCCCAGGGCTGCATCTGCAAAGGGACGTCAGACAAGTGCAGCTGCTGTGCCTGATG 330

Qy        461 TCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGA-------- 512
               | ||||||| |||||  |||| || | ||||  |||| | |||| ||||
Db        331 CCAGGACAGCTGTGCTCTCAGATGTAAATAGAGCAACCTATATAAACCTGGATTTTTTTT 390

Qy        513 -ATTTTTTTTCCATACAACCCTGCCCCATCTACT--GTATTTTTTTTAATGAAATATGTG 569
               ||||||||   |||||||||| ||| | | ||   | |||||||  ||||||||||||
Db        391 TTTTTTTTTTTTGTACAACCCTGACCCGTTTGCTACATCTTTTTTTCTATGAAATATGTG 450

Qy        570 AATGATAATA 579
              ||||  ||||
Db        451 AATGGCAATA 460


RESULT 11
BM921692
LOCUS       BM921692                1074 bp    mRNA    linear   EST 12-MAR-2002
DEFINITION  AGENCOURT_6707922 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5753242
```

```
            5', mRNA sequence.
ACCESSION   BM921692
VERSION     BM921692.1  GI:19372071
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1074)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12788  row: l  column: 11
            High quality sequence stop: 430.
FEATURES             Location/Qualifiers
     source          1. .1074
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5753242"
                     /clone_lib="NIH_MGC_115"
                     /lab_host="DH10B"
                     /note="Organ: pooled brain, lung, testis; Vector:
                     pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
                     source anonymous pool of 6 male brains, age range 23-27; 1
                     male lung, age 27; and 1 male testis, age 69.  Library is
                     oligo-dT primed and directionally cloned (EcoRV site is
                     destroyed upon  cloning).  Average insert size 1.8 kb,
                     insert size range 1-3 kb. Library is normalized and
                     enriched for full-length clones and was  constructed by C.
                     Gruber (Invitrogen).  Research Genetics tracking  code
                     021. Note: this is a NIH_MGC Library."
BASE COUNT      267 a    315 c    210 g    278 t      4 others
ORIGIN

  Query Match             47.5%;  Score 276.8;  DB 14;  Length 1074;
  Best Local Similarity   88.4%;  Pred. No. 8.2e-65;
  Matches  327;  Conservative    0;  Mismatches   32;  Indels   11;  Gaps    2;

Qy        221 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          3 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 62

Qy        281 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         63 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 122

Qy        341 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAA 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAA 182

Qy        401 GTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATG 460
              |||||||||||||||||||||||||| | ||||| ||||||||||||||||||||||||
Db        183 GTGTGCCCAGGGCTGCATCTGCAAAGGGACGTCAGACAAGTGCAGCTGCTGTGCCTGATG 242

Qy        461 TCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGA-------- 512
              | ||||||| ||||| |||| || | |||| |||| | |||| ||||
Db        243 CCAGGACAGCTGTGCTCTCAGATGTAAATAGAGCAACCTATATAAACCTGGATTTTTTTT 302

Qy        513 -ATTTTTTTTCCATACAACCCTGCCCCATCTACT--GTATTTTTTTTAATGAAATATGTG 569
               ||||||||  |||||||||| ||| | | ||   | |||||||  ||||||||||||
Db        303 TTTTTTTTTTTTGTACAACCCTGACCCGTTTGCTACATCTTTTTTTCTATGAAATATGTG 362

Qy        570 AATGATAATA 579
              ||||  ||||
Db        363 AATGGCAATA 372


RESULT 12
BM544354
LOCUS       BM544354               1159 bp    mRNA    linear   EST 20-FEB-2002
DEFINITION  AGENCOURT_6491716 NIH_MGC_124 Homo sapiens cDNA clone IMAGE:5728058
            5', mRNA sequence.
ACCESSION   BM544354
VERSION     BM544354.1  GI:18775545
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1159)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Invitrogen
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12723  row: c  column: 03
            High quality sequence stop: 517.
FEATURES             Location/Qualifiers
     source          1. .1159
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5728058"
                     /clone_lib="NIH_MGC_124"
                     /tissue_type="hippocampus"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: EcoRV
                     (destroyed); Site_2: NotI; RNA source male hippocampus,
                     age 27. Library is oligo-dT primed and directionally
                     cloned (EcoRV site is destroyed upon cloning). Average
                     insert size 1.4 kb, insert size range 0.9-4 kb. Library is
                     normalized and enriched for full-length clones and was
                     constructed by C. Gruber (Invitrogen). Research Genetics
                     tracking code 012."
BASE COUNT      366 a    279 c    247 g    267 t
ORIGIN

  Query Match             47.5%;  Score 276.8;  DB 13;  Length 1159;
  Best Local Similarity   88.4%;  Pred. No. 8.4e-65;
  Matches  327;  Conservative    0;  Mismatches   32;  Indels   11;  Gaps    2;

Qy        221 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          4 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 63

Qy        281 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         64 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 123

Qy        341 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAA 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        124 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAA 183

Qy        401 GTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATG 460
              |||||||||||||||||||||||||| | ||||| ||||||||||||||||||||||||
Db        184 GTGTGCCCAGGGCTGCATCTGCAAAGGGACGTCAGACAAGTGCAGCTGCTGTGCCTGATG 243

Qy        461 TCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGA-------- 512
              | ||||||| ||||| |||| || | |||| |||| | |||| ||||
Db        244 CCAGGACAGCTGTGCTCTCAGATGTAAATAGAGCAACCTATATAAACCTGGATTTTTTTT 303
```

```
Qy        513 -ATTTTTTTTCCATACAACCCTGCCCCATCTACT--GTATTTTTTTTAATGAAATATGTG 569
              ||||||||   |||||||||| ||| | | ||   | |||||||  ||||||||||||
Db        304 TTTTTTTTTTTTGTACAACCCTGACCCGTTTGCTACATCTTTTTTTCTATGAAATATGTG 363

Qy        570 AATGATAATA 579
              ||||  ||||
Db        364 AATGGCAATA 373


RESULT 13
BG897617
LOCUS       BG897617                493 bp    mRNA    linear   EST 06-NOV-2001
DEFINITION  HOA15-1-B10 HOA (Human Osteoarthritic Cartilage) Homo sapiens cDNA,
            mRNA sequence.
ACCESSION   BG897617
VERSION     BG897617.1  GI:14307858
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 493)
  AUTHORS   Kumar,S., Connor,J.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J.,
            Sathe,G., Mui,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and
            Lark,M.W.
  TITLE     Identification and initial characterization of 5000 expressed
            sequenced tags (ESTs) each from adult human normal and
            osteoarthritic cartilage cDNA libraries
  JOURNAL   Osteoarthr. Cartil. 9 (7), 641-653 (2001)
  MEDLINE   21482651
COMMENT     Contact: Sanjay Kumar
            UW2109
            GlaxoSmithKline
            709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA
            Tel: 610-270-7245
            Fax: 610-270-5598
            Email: sanjay_kumar-1@gsk.com
            Seq primer: T7.
FEATURES             Location/Qualifiers
     source          1. .493
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="HOA (Human Osteoarthritic Cartilage)"
                     /tissue_type="cartilage"
                     /lab_host="E.coli DH10 B"
                     /note="Vector: pSPORT I; Site_1: SalI; Site_2: NotI;
                     Directional"
BASE COUNT      127 a    128 c    110 g    128 t
ORIGIN

  Query Match             47.3%;  Score 275.8;  DB 12;  Length 493;
  Best Local Similarity   87.5%;  Pred. No. 1.2e-64;
  Matches  328;  Conservative    0;  Mismatches   37;  Indels   10;  Gaps    2;

Qy        215 CTCGAGCCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAAT 274
              | || | |  ||||||||||||||||||||||||||||||||||||||||||||||||||
Db         75 CACGCGTCCGTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAAT 134

Qy        275 GGACCCCAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATG 334
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        135 GGACCCCAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATG 194

Qy        335 CAAAGAGTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTG 394
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        195 CAAAGAGTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTG 254

Qy        395 TGCCAAGTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGC 454
              |||||||||||||||||||||||||||||| | ||||| ||||||||||||||||||||
Db        255 TGCCAAGTGTGCCCAGGGCTGCATCTGCAAAGGGACGTCAGACAAGTGCAGCTGCTGTGC 314

Qy        455 CTGATGTCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGA-- 512
              |||||| | |||||||  |||||  |||| || | ||||  ||||  | |||| ||||
Db        315 CTGATGCCAGGACAGCTGTGCTCTCAGATGTAAATAGAGCAACCTATATAAACCTGGATT 374
```

```
Qy        513 ------ATTTTTTTTCCATACAACCCTGCCCCATCTACT--GTATTTTTTTTAATGAAAT 564
                   ||||||||   |||||||||| ||| | | ||   | |||||||  |||||||
Db        375 TTTTTTTTTTTTTTTTTGTACAACCCTGACCCGTTTGCTACATCTTTTTTTCTATGAAAT 434

Qy        565 ATGTGAATGATAATA 579
              |||||||||  ||||
Db        435 ATGTGAATGGCAATA 449


RESULT 14
BG925124
LOCUS       BG925124                576 bp    mRNA    linear   EST 06-NOV-2001
DEFINITION  HNC37-1-B5.R HNC (Human Normal Cartilage) Homo sapiens cDNA, mRNA
            sequence.
ACCESSION   BG925124
VERSION     BG925124.1  GI:14319647
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 576)
  AUTHORS   Kumar,S., Connor,J.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J.,
            Sathe,G., Mui,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and
            Lark,M.W.
  TITLE     Identification and initial characterization of 5000 expressed
            sequenced tags (ESTs) each from adult human normal and
            osteoarthritic cartilage cDNA libraries
  JOURNAL   Osteoarthr. Cartil. 9 (7), 641-653 (2001)
  MEDLINE   21482651
COMMENT     Contact: Sanjay Kumar
            UW2109
            GlaxoSmithKline
            709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA
            Tel: 610-270-7245
            Fax: 610-270-5598
            Email: sanjay_kumar-1@gsk.com
            Seq primer: T7.
FEATURES             Location/Qualifiers
     source          1. .576
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="HNC (Human Normal Cartilage)"
                     /tissue_type="cartilage"
                     /lab_host="E.coli DH10 B"
                     /note="Vector: pSPORT I; Site_1: SalI; Site_2: NotI;
                     Directional"
BASE COUNT      150 a    150 c    129 g    147 t
ORIGIN

  Query Match             47.2%;  Score 275.2;  DB 13;  Length 576;
  Best Local Similarity   88.1%;  Pred. No. 1.9e-64;
  Matches  326;  Conservative    0;  Mismatches   33;  Indels   11;  Gaps    2;

Qy        221 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         90 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 149

Qy        281 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        150 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 209

Qy        341 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAA 400
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        210 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTTTGGGCTGTGCCAA 269

Qy        401 GTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATG 460
              |||||||||||||||||||||||||||| | ||||| ||||||||||||||||||||||
Db        270 GTGTGCCCAGGGCTGCATCTGCAAAGGGACGTCAGACAAGTGCAGCTGCTGTGCCTGATG 329

Qy        461 TCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGA-------- 512
               | |||||||  |||||  |||| || | ||||  ||||  | |||| ||||
```

```
Db        330 CCAGGACAGCTGTGCTCTCAGATGTAAATAGAGCAACCTATATAAACCTGGATTTTTTTT 389

Qy        513 -ATTTTTTTTCCATACAACCCTGCCCCATCTACT--GTATTTTTTTTAATGAAATATGTG 569
              ||||||||   |||||||||| ||| | | ||   | |||||||  ||||||||||||
Db        390 TTTTTTTTTTTTGTACAACCCTGACCCGTTTGCTACATCTTTTTTTCTATGAAATATGTG 449

Qy        570 AATGATAATA 579
              ||||  ||||
Db        450 AATGGCAATA 459


RESULT 15
AV661102
LOCUS       AV661102                 407 bp    mRNA    linear   EST 16-JAN-2002
DEFINITION  AV661102 GLC Homo sapiens cDNA clone GLCGOF09 3', mRNA sequence.
ACCESSION   AV661102
VERSION     AV661102.1  GI:9882116
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 407)
  AUTHORS   Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
            Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
            Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
            Hu,G., Gu,J., Chen,Z. and Han,Z.
  TITLE     Insight into hepatocellular carcinogenesis at transcriptome level
            by comparing gene expression profiles of hepatocellular carcinoma
            with those of corresponding noncancerous liver
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
  MEDLINE   21625106
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1. .407
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="GLCGOF09"
                     /clone_lib="GLC"
                     /tissue_type="corresponding non cancerous liver tissue"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT       96 a    103 c     89 g    119 t
ORIGIN

  Query Match             47.1%;  Score 274.8;  DB 10;  Length 407;
  Best Local Similarity   88.3%;  Pred. No. 2.2e-64;
  Matches  325;  Conservative    0;  Mismatches   32;  Indels   11;  Gaps    2;

Qy        223 GCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCCCA 282
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCCCA 60

Qy        283 ACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGAGT 342
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGAGT 120

Qy        343 GCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAAGT 402
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAAGT 180

Qy        403 GTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATGTC 462
              |||||||||||||||||||||||| | ||||| ||||||||||||||||||||||| |
Db        181 GTGCCCAGGGCTGCATCTGCAAAGGGACGTCAGACAAGTGCAGCTGCTGTGCCTGATGCC 240

Qy        463 CGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGA---------A 513
              |||||||  |||||  |||| || | ||||  ||||  | |||| ||||
Db        241 AGGACAGCTGTGCTCTCAGATGTAAATAGAGCAACCTATATAAACCTGGATTTTTTTTTT 300

Qy        514 TTTTTTTTCCATACAACCCTGCCCCATCTACT--GTATTTTTTTTAATGAAATATGTGAA 571
              ||||||||   |||||||||| ||| | | ||   | |||||||  ||||||||||||||
Db        301 TTTTTTTTTTGTACAACCCTGACCCGTTTGCTACATCTTTTTTTCTATGAAATATGTGAA 360

Qy        572 TGATAATA 579
              ||  ||||
Db        361 TGGCAATA 368


Search completed: July 16, 2003, 14:55:43
Job time : 640.456 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 16, 2003, 10:35:27 ; Search time 22.7443 Seconds
(without alignments)
7860.974 Million cell updates/sec

Title: US-09-981-353-114
Perfect score: 583
Sequence: 1 ccctgtaggtgctgtccttg.....tatgtgaatgataatagtca 583

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	216	37.0	958	3	US-09-049-672A-15	Sequence 15, Appl
c	2	214.4	36.8	3819	4	US-09-042-353-393	Sequence 393, App
c	3	214.4	36.8	3819	4	US-08-758-417A-243	Sequence 243, App
c	4	212.8	36.5	2178	1	US-08-463-587A-24	Sequence 24, Appl
c	5	212.8	36.5	2178	2	US-08-463-667A-2	Sequence 2, Appli
c	6	212.8	36.5	2178	3	US-08-923-854-24	Sequence 24, Appl
c	7	212.8	36.5	2178	5	PCT-US91-09133-25	Sequence 25, Appl
c	8	212.8	36.5	7305	1	US-08-286-740-4	Sequence 4, Appli
c	9	212.8	36.5	7305	5	PCT-US95-09576-4	Sequence 4, Appli
c	10	212	36.4	468	1	US-08-236-311-11	Sequence 11, Appl
c	11	212	36.4	468	3	US-08-457-918-11	Sequence 11, Appl
c	12	212	36.4	642	1	US-08-157-101A-8	Sequence 8, Appli
c	13	212	36.4	1066	1	US-08-157-101A-4	Sequence 4, Appli
c	14	211.2	36.2	705	4	US-09-171-945-51	Sequence 51, Appl
c	15	211.2	36.2	705	4	US-09-171-945-96	Sequence 96, Appl
c	16	211.2	36.2	705	4	US-09-171-945-98	Sequence 98, Appl
c	17	211.2	36.2	708	1	US-08-488-376-18	Sequence 18, Appl
c	18	211.2	36.2	708	2	US-08-634-223-18	Sequence 18, Appl
c	19	211.2	36.2	708	2	US-08-634-224-18	Sequence 18, Appl
c	20	211.2	36.2	708	2	US-08-634-400-18	Sequence 18, Appl
c	21	211.2	36.2	708	2	US-08-635-878-18	Sequence 18, Appl
c	22	211.2	36.2	708	2	US-08-770-057-18	Sequence 18, Appl
c	23	211.2	36.2	708	4	US-09-335-697B-18	Sequence 18, Appl
c	24	211.2	36.2	708	4	US-09-335-697B-18	Sequence 18, Appl
c	25	211.2	36.2	780	3	US-09-027-449-54	Sequence 54, Appl
c	26	211.2	36.2	780	3	US-09-027-449-58	Sequence 58, Appl
c	27	211.2	36.2	780	3	US-09-027-449-65	Sequence 65, Appl
c	28	211.2	36.2	780	3	US-08-804-444A-54	Sequence 54, Appl
c	29	211.2	36.2	780	3	US-08-804-444A-58	Sequence 58, Appl
c	30	211.2	36.2	780	3	US-09-026-985-54	Sequence 54, Appl
c	31	211.2	36.2	780	3	US-09-026-985-58	Sequence 58, Appl
c	32	211.2	36.2	780	3	US-09-026-985-65	Sequence 65, Appl
c	33	211.2	36.2	780	4	US-09-121-952A-54	Sequence 54, Appl
c	34	211.2	36.2	780	4	US-09-121-952A-58	Sequence 58, Appl
c	35	211.2	36.2	780	4	US-09-121-952A-65	Sequence 65, Appl
c	36	211.2	36.2	780	4	US-09-234-340A-54	Sequence 54, Appl
c	37	211.2	36.2	780	4	US-09-234-340A-58	Sequence 58, Appl
c	38	211.2	36.2	780	4	US-09-234-340A-65	Sequence 65, Appl
c	39	211.2	36.2	1576	5	PCT-US95-11405-34	Sequence 34, Appl
c	40	211.2	36.2	2143	3	US-09-097-309-5	Sequence 5, Appli
c	41	211.2	36.2	2143	4	US-09-097-171A-9	Sequence 9, Appli
c	42	211.2	36.2	2143	4	US-09-460-587-5	Sequence 5, Appli
c	43	211.2	36.2	6127	2	US-08-887-352B-1	Sequence 1, Appli
c	44	211.2	36.2	6127	4	US-09-109-207C-1	Sequence 1, Appli
c	45	211.2	36.2	6127	4	US-09-296-005-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT 1
US-09-049-672A-15/c
; Sequence 15, Application US/09049672A
; Patent No. 6135941
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Tang, Y. Tom
;    APPLICANT:  Yue, Henry
;    APPLICANT:  Au-Young, Janice
;    APPLICANT:  Corley, Neil C.
;    APPLICANT:  Guegler, Karl J.
;    APPLICANT:  Baughn, Mariah R.
;    TITLE OF INVENTION:  HUMAN IMMUNE SYSTEM ASSOCIATED PROTEINS
;    NUMBER OF SEQUENCES:  28
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/049,672A
;      FILING DATE:  HEREWITH
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cerrone, Michael C
;      REGISTRATION NUMBER:  39,132
;      REFERENCE/DOCKET NUMBER:  PF-0497 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-855-0555
;      TELEFAX:  650-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  ADENINB01
```

```
;      CLONE:  161752
US-09-049-672A-15

  Query Match             37.0%;  Score 216;  DB 3;  Length 958;
  Best Local Similarity   97.8%;  Pred. No. 2.2e-56;
  Matches  219;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          3 CTGTAGGTGCTGTCCTTGCTGTCCTGCTCTGTGACACTCTCCTGGGAGTCACCCGATTGG 62
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        623 CTGTAGGTGCTGTCCTTGCTGTCCTGCTCTGTGACACTCTCCTGGGAGTTACCCGATTGG 564

Qy         63 AGGGCGTTATCCACCTTCCACTGTACTTTGGCCTCTCTGGGATAGAAGTTATTCAGCAGG 122
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        563 AGGGCGTTATCCACCTTCCACTGTACTTTGGCCTCTCTGGGATAGAAGTTATTCAGCAGG 504

Qy        123 CACACAACAGAGGCAGTTCCAGATTTCAACTGCTCATCAGATGGCGGGAAGATGAAGACA 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        503 CACACAACAGAGGCAGTTCCAGATTTCAACTGCTCATCAGATGGCGGGAAGATGAAGACA 444

Qy        183 GATGGTGCAGCCACAGTTCGTTTGATCTCCAGCTCGAGCCGCTG 226
              |||||||||||||||||||||||||||||||| |  || |||
Db        443 GATGGTGCAGCCACAGTTCGTTTGATCTCCAGCTTGGTCCCCTG 400


RESULT 2
US-09-042-353-393/c
; Sequence 393, Application US/09042353
; Patent No. 6255458
;  GENERAL INFORMATION:
;    APPLICANT:  Lonberg, Nils
;    APPLICANT:  Kay, Robert M.
;    TITLE OF INVENTION:  Transgenic No. 6255458-Human Animals for
;    TITLE OF INVENTION:  Producing Heterologous Antibodies
;    NUMBER OF SEQUENCES:  421
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew LLP
;      STREET:  Two Embarcadero Center, Eighth Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/042,353
;      FILING DATE:  13-MAR-1998
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/810,279
;      FILING DATE:  17-DEC-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/853,408
;      FILING DATE:  18-MAR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/904,068
;      FILING DATE:  23-JUN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/990,860
;      FILING DATE:  16-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/053,131
;      FILING DATE:  26-APR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/096,762
;      FILING DATE:  22-JUL-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/155,301
;      FILING DATE:  18-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,739
;      FILING DATE:  03-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/165,699
;      FILING DATE:  10-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/209,741
;      FILING DATE:  09-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/352,322
;      FILING DATE:  07-DEC-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/544,404
;      FILING DATE:  10-OCT-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/728,463
;      FILING DATE:  10-OCT-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/US96/16433
;      FILING DATE:  10-OCT-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/758,417
;      FILING DATE:  02-DEC-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/US97/21803
;      FILING DATE:  01-DEC-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Apple, Randolph T.
;      REGISTRATION NUMBER:  36,429
;      REFERENCE/DOCKET NUMBER:  014643-009040US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 576-0200
;      TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  393:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3819 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-09-042-353-393

  Query Match             36.8%;  Score 214.4;  DB 4;  Length 3819;
  Best Local Similarity   97.3%;  Pred. No. 1.4e-55;
  Matches  218;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          3 CTGTAGGTGCTGTCCTTGCTGTCCTGCTCTGTGACACTCTCCTGGGAGTCACCCGATTGG 62
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db       3026 CTGTAGGTGCTGTCCTTGCTGTCCTGCTCTGTGACACTCTCCTGGGAGTTACCCGATTGG 2967

Qy         63 AGGGCGTTATCCACCTTCCACTGTACTTTGGCCTCTCTGGGATAGAAGTTATTCAGCAGG 122
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2966 AGGGCGTTATCCACCTTCCACTGTACTTTGGCCTCTCTGGGATAGAAGTTATTCAGCAGG 2907

Qy        123 CACACAACAGAGGCAGTTCCAGATTTCAACTGCTCATCAGATGGCGGGAAGATGAAGACA 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2906 CACACAACAGAGGCAGTTCCAGATTTCAACTGCTCATCAGATGGCGGGAAGATGAAGACA 2847

Qy        183 GATGGTGCAGCCACAGTTCGTTTGATCTCCAGCTCGAGCCGCTG 226
              |||||||||||||||||||||||||||||||| |   | |||
Db       2846 GATGGTGCAGCCACAGTTCGTTTGATCTCCAGCTTGGTTCCCTG 2803


RESULT 3
US-08-758-417A-243/c
; Sequence 243, Application US/08758417A
; Patent No. 6300129
;  GENERAL INFORMATION:
;       APPLICANT: Lonberg, Nils
;                  Kay, Robert M.
;       TITLE OF INVENTION: Transgenic No. 6300129-Human Animals for
;                           Producing Heterologous Antibodies
;       NUMBER OF SEQUENCES: 417
;       CORRESPONDENCE ADDRESS:
```

```
;           ADDRESSEE: Townsend and Townsend and Crew LLP
;           STREET: Two Embarcadero Center, Eighth Floor
;           CITY: San Francisco
;           STATE: California
;           COUNTRY: USA
;           ZIP: 94111-3834
;         COMPUTER READABLE FORM:
;           MEDIUM TYPE: Floppy disk
;           COMPUTER: IBM PC compatible
;           OPERATING SYSTEM: PC-DOS/MS-DOS
;           SOFTWARE: PatentIn Release #1.0, Version #1.30
;         CURRENT APPLICATION DATA:
;           APPLICATION NUMBER: US/08/758,417A
;           FILING DATE: 02-Dec-1996
;           CLASSIFICATION: <Unknown>
;         PRIOR APPLICATION DATA:
;           APPLICATION NUMBER: US 08/728,463
;           FILING DATE: 10-OCT-1996
;           APPLICATION NUMBER: US 08/544,404
;           FILING DATE: 10-OCT-1995
;           APPLICATION NUMBER: US 08/352,322
;           FILING DATE: 07-DEC-1994
;           APPLICATION NUMBER: US 08/209,741
;           FILING DATE: 09-MAR-1994
;           APPLICATION NUMBER: US 08/165,699
;           FILING DATE: 10-DEC-1993
;           APPLICATION NUMBER: US 08/161,739
;           FILING DATE: 03-DEC-1993
;           APPLICATION NUMBER: US 08/155,301
;           FILING DATE: 18-NOV-1993
;           APPLICATION NUMBER: US 08/096,762
;           FILING DATE: 22-JUL-1993
;           APPLICATION NUMBER: US 08/053,131
;           FILING DATE: 26-APR-1993
;           APPLICATION NUMBER: US 07/990,860
;           FILING DATE: 16-DEC-1992
;         ATTORNEY/AGENT INFORMATION:
;           NAME: Serafini, Andrew T.
;           REGISTRATION NUMBER: 41,303
;           REFERENCE/DOCKET NUMBER: 014643-009030US
;         TELECOMMUNICATION INFORMATION:
;           TELEPHONE: (415) 576-0200
;           TELEFAX: (415) 576-0300
;       INFORMATION FOR SEQ ID NO: 243:
;         SEQUENCE CHARACTERISTICS:
;           LENGTH: 3819 base pairs
;           TYPE: nucleic acid
;           STRANDEDNESS: single
;           TOPOLOGY: linear
;         MOLECULE TYPE: DNA
;         SEQUENCE DESCRIPTION: SEQ ID NO: 243:
US-08-758-417A-243

  Query Match           36.8%;  Score 214.4;  DB 4;  Length 3819;
  Best Local Similarity  97.3%;  Pred. No. 1.4e-55;
  Matches  218;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          3 CTGTAGGTGCTGTCCTTGCTGTCCTGCTCTGTGACACTCTCCTGGGAGTCACCCGATTGG 62
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db       3026 CTGTAGGTGCTGTCCTTGCTGTCCTGCTCTGTGACACTCTCCTGGGAGTTACCCGATTGG 2967

Qy         63 AGGGCGTTATCCACCTTCCACTGTACTTTGGCCTCTCTGGGATAGAAGTTATTCAGCAGG 122
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2966 AGGGCGTTATCCACCTTCCACTGTACTTTGGCCTCTCTGGGATAGAAGTTATTCAGCAGG 2907

Qy        123 CACACAACAGAGGCAGTTCCAGATTTCAACTGCTCATCAGATGGCGGGAAGATGAAGACA 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2906 CACACAACAGAGGCAGTTCCAGATTTCAACTGCTCATCAGATGGCGGGAAGATGAAGACA 2847

Qy        183 GATGGTGCAGCCACAGTTCGTTTGATCTCCAGCTCGAGCCGCTG 226
              |||||||||||||||||||||||||||||||| |   | |||
Db       2846 GATGGTGCAGCCACAGTTCGTTTGATCTCCAGCTTGGTTCCCTG 2803

RESULT 4
US-08-463-587A-24/c
; Sequence 24, Application US/08463587A
; Patent No. 5821047
; GENERAL INFORMATION:
;   APPLICANT:  Garrard, Lisa J.
;   APPLICANT:  Henner, Dennis J.
;   APPLICANT:  Bass, Steven
;   APPLICANT:  Greene, Ronald
;   APPLICANT:  Lowman, Henry B.
;   APPLICANT:  Wells, James A.
;   APPLICANT:  Matthews, David J.
;   TITLE OF INVENTION:  ENRICHMENT METHOD FOR VARIANT PROTEINS WITH
;   TITLE OF INVENTION:  ALTERED BINDING PROPERTIES
;   NUMBER OF SEQUENCES:  43
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Genentech, Inc.
;     STREET:  460 Point San Bruno Blvd
;     CITY:  South San Francisco
;     STATE:  California
;     COUNTRY:  USA
;     ZIP:  94080
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  WinPatin (Genentech)
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/463,587A
;     FILING DATE:  05-Jun-1995
;     CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  08/050058
;     FILING DATE:  30-APR-1993
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  PCT/US91/09133
;     FILING DATE:  03-DEC-1991
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  07/743614
;     FILING DATE:  09-AUG-1991
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  07/715300
;     FILING DATE:  14-JUN-1991
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  07/683400
;     FILING DATE:  10-APR-1991
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  07/621667
;     FILING DATE:  03-DEC-1990
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Schwartz, Timothy R.
;     REGISTRATION NUMBER:  32171
;     REFERENCE/DOCKET NUMBER:  P0645P4D2
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  415/225-7467
;     TELEFAX:  415/952-9881
;     TELEX:  910/371-7168
; INFORMATION FOR SEQ ID NO:  24:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  2178 base pairs
;     TYPE:  Nucleic Acid
;     STRANDEDNESS:  Single
;     TOPOLOGY:  Linear
US-08-463-587A-24

  Query Match           36.5%;  Score 212.8;  DB 1;  Length 2178;
  Best Local Similarity  96.9%;  Pred. No. 3.2e-55;
  Matches  217;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          3 CTGTAGGTGCTGTCCTTGCTGTCCTGCTCTGTGACACTCTCCTGGGAGTCACCCGATTGG 62
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        590 CTGTAGGTGCTGTCCTTGCTGTCCTGCTCTGTGACACTCTCCTGGGAGTTACCCGATTGG 531
```

```
Qy         63 AGGGCGTTATCCACCTTCCACTGTACTTTGGCCTCTCTGGGATAGAAGTTATTCAGCAGG 122
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        530 AGGGCGTTATCCACCTTCCACTGTACTTTGGCCTCTCTGGGATAGAAGTTATTCAGCAGG 471

Qy        123 CACACAACAGAGGCAGTTCCAGATTTCAACTGCTCATCAGATGGCGGGAAGATGAAGACA 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        470 CACACAACAGAGGCAGTTCCAGATTTCAACTGCTCATCAGATGGCGGGAAGATGAAGACA 411

Qy        183 GATGGTGCAGCCACAGTTCGTTTGATCTCCAGCTCGAGCCGCTG 226
              |||||||||||||||||||||||||||||| || |   | |||
Db        410 GATGGTGCAGCCACAGTTCGTTTGATCTCCACCTTGGTACCCTG 367


RESULT 5
US-08-463-667A-2/c
; Sequence 2, Application US/08463667A
; Patent No. 5834598
;  GENERAL INFORMATION:
;    APPLICANT:  Gerrard, Lisa J.
;    APPLICANT:  Henner, Dennis J.
;    APPLICANT:  Bass, Steven
;    APPLICANT:  Greene, Ronald
;    APPLICANT:  Lowman, Henry B.
;    APPLICANT:  Wells, James A.
;    APPLICANT:  Matthews, David J.
;    TITLE OF INVENTION:  ENRICHMENT METHOD FOR VARIANT PROTEINS
;    TITLE OF INVENTION:  WITH ALTERED BINDING PROPERTIES
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/463,667A
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/050,058
;      FILING DATE:  30-APR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/US91/09133
;      FILING DATE:  03-DEC-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/743,614
;      FILING DATE:  09-AUG-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/715,300
;      FILING DATE:  14-JUN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/683,400
;      FILING DATE:  10-APR-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/621,667
;      FILING DATE:  03-DEC-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Winter, Daryl B.
;      REGISTRATION NUMBER:  32,637
;      REFERENCE/DOCKET NUMBER:  645P4D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 225-1249
;      TELEFAX:  (415) 952-9881
;      TELEX:  910 371-7168
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2178 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..711
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  796..2178
US-08-463-667A-2

  Query Match             36.5%;  Score 212.8;  DB 2;  Length 2178;
  Best Local Similarity   96.9%;  Pred. No. 3.2e-55;
  Matches  217;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          3 CTGTAGGTGCTGTCCTTGCTGTCCTGCTCTGTGACACTCTCCTGGGAGTCACCCGATTGG 62
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        590 CTGTAGGTGCTGTCCTTGCTGTCCTGCTCTGTGACACTCTCCTGGGAGTTACCCGATTGG 531

Qy         63 AGGGCGTTATCCACCTTCCACTGTACTTTGGCCTCTCTGGGATAGAAGTTATTCAGCAGG 122
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        530 AGGGCGTTATCCACCTTCCACTGTACTTTGGCCTCTCTGGGATAGAAGTTATTCAGCAGG 471

Qy        123 CACACAACAGAGGCAGTTCCAGATTTCAACTGCTCATCAGATGGCGGGAAGATGAAGACA 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        470 CACACAACAGAGGCAGTTCCAGATTTCAACTGCTCATCAGATGGCGGGAAGATGAAGACA 411

Qy        183 GATGGTGCAGCCACAGTTCGTTTGATCTCCAGCTCGAGCCGCTG 226
              |||||||||||||||||||||||||||||| || |   | |||
Db        410 GATGGTGCAGCCACAGTTCGTTTGATCTCCACCTTGGTACCCTG 367


RESULT 6
US-08-923-854-24/c
; Sequence 24, Application US/08923854
; Patent No. 6040136
;  GENERAL INFORMATION:
;    APPLICANT:  Garrard, Lisa J.
;    APPLICANT:  Henner, Dennis J.
;    APPLICANT:  Bass, Steven
;    APPLICANT:  Greene, Ronald
;    APPLICANT:  Lowman, Henry B.
;    APPLICANT:  Wells, James A.
;    APPLICANT:  Matthews, David J.
;    TITLE OF INVENTION:  ENRICHMENT METHOD FOR VARIANT PROTEINS WITH
;    TITLE OF INVENTION:  ALTERED BINDING PROPERTIES
;    NUMBER OF SEQUENCES:  43
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/923,854
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/463587
;      FILING DATE:  05-Jun-1995
;      APPLICATION NUMBER:  08/050058
;      FILING DATE:  30-APR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US91/09133
;      FILING DATE:  03-DEC-1991
```

```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/743614
;      FILING DATE:  09-AUG-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/715300
;      FILING DATE:  14-JUN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/683400
;      FILING DATE:  10-APR-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/621667
;      FILING DATE:  03-DEC-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Schwartz, Timothy R.
;      REGISTRATION NUMBER:  32171
;      REFERENCE/DOCKET NUMBER:  P0645P4D2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-7467
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2178 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-923-854-24

  Query Match             36.5%;  Score 212.8;  DB 3;  Length 2178;
  Best Local Similarity   96.9%;  Pred. No. 3.2e-55;
  Matches  217;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          3 CTGTAGGTGCTGTCCTTGCTGTCCTGCTCTGTGACACTCTCCTGGGAGTCACCCGATTGG 62
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        590 CTGTAGGTGCTGTCCTTGCTGTCCTGCTCTGTGACACTCTCCTGGGAGTTACCCGATTGG 531

Qy         63 AGGGCGTTATCCACCTTCCACTGTACTTTGGCCTCTCTGGGATAGAAGTTATTCAGCAGG 122
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        530 AGGGCGTTATCCACCTTCCACTGTACTTTGGCCTCTCTGGGATAGAAGTTATTCAGCAGG 471

Qy        123 CACACAACAGAGGCAGTTCCAGATTTCAACTGCTCATCAGATGGCGGGAAGATGAAGACA 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        470 CACACAACAGAGGCAGTTCCAGATTTCAACTGCTCATCAGATGGCGGGAAGATGAAGACA 411

Qy        183 GATGGTGCAGCCACAGTTCGTTTGATCTCCAGCTCGAGCCGCTG 226
              |||||||||||||||||||||||||||||| || |   | |||
Db        410 GATGGTGCAGCCACAGTTCGTTTGATCTCCACCTTGGTACCCTG 367


RESULT 7
PCT-US91-09133-25/c
; Sequence 25, Application PC/TUS9109133
;  GENERAL INFORMATION:
;    APPLICANT:  Genentech, Inc.
;    APPLICANT:  Garrard, Lisa J.
;    APPLICANT:  Henner, Dennis J.
;    APPLICANT:  Bass, Steven
;    APPLICANT:  Greene, Ronald
;    APPLICANT:  Lowman, Henry B.
;    APPLICANT:  Wells, James A.
;    APPLICANT:  Matthews, David J.
;    TITLE OF INVENTION:  Enrichment Method For Variant Proteins
;    TITLE OF INVENTION:  With Altered Binding Properties
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US91/09133
;      FILING DATE:  19911203
;      CLASSIFICATION:  425
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/743614
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/715300
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/683400
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/621667
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Benson, Robert H.
;      REGISTRATION NUMBER:  30,446
;      REFERENCE/DOCKET NUMBER:  645P4
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/266-1489
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2178 bases
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
PCT-US91-09133-25

  Query Match             36.5%;  Score 212.8;  DB 5;  Length 2178;
  Best Local Similarity   96.9%;  Pred. No. 3.2e-55;
  Matches  217;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          3 CTGTAGGTGCTGTCCTTGCTGTCCTGCTCTGTGACACTCTCCTGGGAGTCACCCGATTGG 62
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        590 CTGTAGGTGCTGTCCTTGCTGTCCTGCTCTGTGACACTCTCCTGGGAGTTACCCGATTGG 531

Qy         63 AGGGCGTTATCCACCTTCCACTGTACTTTGGCCTCTCTGGGATAGAAGTTATTCAGCAGG 122
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        530 AGGGCGTTATCCACCTTCCACTGTACTTTGGCCTCTCTGGGATAGAAGTTATTCAGCAGG 471

Qy        123 CACACAACAGAGGCAGTTCCAGATTTCAACTGCTCATCAGATGGCGGGAAGATGAAGACA 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        470 CACACAACAGAGGCAGTTCCAGATTTCAACTGCTCATCAGATGGCGGGAAGATGAAGACA 411

Qy        183 GATGGTGCAGCCACAGTTCGTTTGATCTCCAGCTCGAGCCGCTG 226
              |||||||||||||||||||||||||||||| || |   | |||
Db        410 GATGGTGCAGCCACAGTTCGTTTGATCTCCACCTTGGTACCCTG 367


RESULT 8
US-08-286-740-4/c
; Sequence 4, Application US/08286740
; Patent No. 5561053
;  GENERAL INFORMATION:
;    APPLICANT:  Crowley, Craig W.
;    TITLE OF INVENTION:  METHOD FOR SELECTING HIGH-EXPRESSING
;    TITLE OF INVENTION:  HOST CELLS
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/286,740
;      FILING DATE:  05-AUG-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lee, Wendy M.
;      REGISTRATION NUMBER:
;      REFERENCE/DOCKET NUMBER:  798
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1994
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7305 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-286-740-4

  Query Match           36.5%;  Score 212.8;  DB 1;  Length 7305;
  Best Local Similarity  96.9%;  Pred. No. 6.1e-55;
  Matches  217;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          3 CTGTAGGTGCTGTCCTTGCTGTCCTGCTCTGTGACACTCTCCTGGGAGTCACCCGATTGG 62
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db       3416 CTGTAGGTGCTGTCCTTGCTGTCCTGCTCTGTGACACTCTCCTGGGAGTTACCCGATTGG 3357

Qy         63 AGGGCGTTATCCACCTTCCACTGTACTTTGGCCTCTCTGGGATAGAAGTTATTCAGCAGG 122
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3356 AGGGCGTTATCCACCTTCCACTGTACTTTGGCCTCTCTGGGATAGAAGTTATTCAGCAGG 3297

Qy        123 CACACAACAGAGGCAGTTCCAGATTTCAACTGCTCATCAGATGGCGGGAAGATGAAGACA 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3296 CACACAACAGAGGCAGTTCCAGATTTCAACTGCTCATCAGATGGCGGGAAGATGAAGACA 3237

Qy        183 GATGGTGCAGCCACAGTTCGTTTGATCTCCAGCTCGAGCCGCTG 226
              |||||||||||||||||||||||||||||| || |   | |||
Db       3236 GATGGTGCAGCCACAGTTCGTTTGATCTCCACCTTGGTACCCTG 3193


RESULT 9
PCT-US95-09576-4/c
; Sequence 4, Application PC/TUS9509576
;  GENERAL INFORMATION:
;    APPLICANT:  GENENTECH, INC.
;    TITLE OF INVENTION:  METHOD FOR SELECTING HIGH-EXPRESSING
;    TITLE OF INVENTION:  HOST CELLS
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/09576
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/286740
;      FILING DATE:  05-AUG-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lee, Wendy M.
;      REGISTRATION NUMBER:  00,000
;      REFERENCE/DOCKET NUMBER:  798PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1994
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7305 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
PCT-US95-09576-4

  Query Match           36.5%;  Score 212.8;  DB 5;  Length 7305;
  Best Local Similarity  96.9%;  Pred. No. 6.1e-55;
  Matches  217;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          3 CTGTAGGTGCTGTCCTTGCTGTCCTGCTCTGTGACACTCTCCTGGGAGTCACCCGATTGG 62
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db       3416 CTGTAGGTGCTGTCCTTGCTGTCCTGCTCTGTGACACTCTCCTGGGAGTTACCCGATTGG 3357

Qy         63 AGGGCGTTATCCACCTTCCACTGTACTTTGGCCTCTCTGGGATAGAAGTTATTCAGCAGG 122
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3356 AGGGCGTTATCCACCTTCCACTGTACTTTGGCCTCTCTGGGATAGAAGTTATTCAGCAGG 3297

Qy        123 CACACAACAGAGGCAGTTCCAGATTTCAACTGCTCATCAGATGGCGGGAAGATGAAGACA 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3296 CACACAACAGAGGCAGTTCCAGATTTCAACTGCTCATCAGATGGCGGGAAGATGAAGACA 3237

Qy        183 GATGGTGCAGCCACAGTTCGTTTGATCTCCAGCTCGAGCCGCTG 226
              |||||||||||||||||||||||||||||| || |   | |||
Db       3236 GATGGTGCAGCCACAGTTCGTTTGATCTCCACCTTGGTACCCTG 3193


RESULT 10
US-08-236-311-11/c
; Sequence 11, Application US/08236311
; Patent No. 5565335
;  GENERAL INFORMATION:
;    APPLICANT:  Capon, Daniel J.
;    APPLICANT:  Gregory, Timothy J.
;    TITLE OF INVENTION:  Adheson Variants
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/236,311
;      FILING DATE:  02-MAY-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/936190
;      FILING DATE:  26-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/842777
;      FILING DATE:  18-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/250785
;      FILING DATE:  28-SEP-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/104329
;      FILING DATE:  02-OCT-1987
;    ATTORNEY/AGENT INFORMATION:
```

```
;       NAME:  Hasak, Janet E.
;       REGISTRATION NUMBER:  28,616
;       REFERENCE/DOCKET NUMBER:  444P1C2
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  415/225-1896
;       TELEFAX:  415/952-9881
;       TELEX:  910/371-7168
;   INFORMATION FOR SEQ ID NO:  11:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  468 bases
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-236-311-11

  Query Match            36.4%;  Score 212;  DB 1;  Length 468;
  Best Local Similarity   97.7%;  Pred. No. 2.5e-55;
  Matches  215;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          3 CTGTAGGTGCTGTCCTTGCTGTCCTGCTCTGTGACACTCTCCTGGGAGTCACCCGATTGG 62
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        314 CTGTAGGTGCTGTCCTTGCTGTCCTGCTCTGTGACACTCTCCTGGGAGTTACCCGATTGG 255

Qy         63 AGGGCGTTATCCACCTTCCACTGTACTTTGGCCTCTCTGGGATAGAAGTTATTCAGCAGG 122
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        254 AGGGCGTTATCCACCTTCCACTGTACTTTGGCCTCTCTGGGATAGAAGTTATTCAGCAGG 195

Qy        123 CACACAACAGAGGCAGTTCCAGATTTCAACTGCTCATCAGATGGCGGGAAGATGAAGACA 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        194 CACACAACAGAGGCAGTTCCAGATTTCAACTGCTCATCAGATGGCGGGAAGATGAAGACA 135

Qy        183 GATGGTGCAGCCACAGTTCGTTTGATCTCCAGCTCGAGCC 222
              ||||||||||||||||||||||||||||||| || |  ||
Db        134 GATGGTGCAGCCACAGTTCGTTTGATCTCCACCTTGGTCC 95


RESULT 11
US-08-457-918-11/c
; Sequence 11, Application US/08457918
; Patent No. 6117655
;  GENERAL INFORMATION:
;    APPLICANT:  Capon, Daniel J.
;    APPLICANT:  Gregory, Timothy J.
;    TITLE OF INVENTION:  Adheson Variants
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/457,918
;      FILING DATE:  1-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/236311
;      FILING DATE:  02-MAY-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/936190
;      FILING DATE:  26-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/842777
;      FILING DATE:  18-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/250785
;      FILING DATE:  28-SEP-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/104329
;      FILING DATE:  02-OCT-1987
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kubinec, Jeffrey S.
;      REGISTRATION NUMBER:  36,575
;      REFERENCE/DOCKET NUMBER:  P0444P1C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-8228
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  468 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-457-918-11

  Query Match            36.4%;  Score 212;  DB 3;  Length 468;
  Best Local Similarity   97.7%;  Pred. No. 2.5e-55;
  Matches  215;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          3 CTGTAGGTGCTGTCCTTGCTGTCCTGCTCTGTGACACTCTCCTGGGAGTCACCCGATTGG 62
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        314 CTGTAGGTGCTGTCCTTGCTGTCCTGCTCTGTGACACTCTCCTGGGAGTTACCCGATTGG 255

Qy         63 AGGGCGTTATCCACCTTCCACTGTACTTTGGCCTCTCTGGGATAGAAGTTATTCAGCAGG 122
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        254 AGGGCGTTATCCACCTTCCACTGTACTTTGGCCTCTCTGGGATAGAAGTTATTCAGCAGG 195

Qy        123 CACACAACAGAGGCAGTTCCAGATTTCAACTGCTCATCAGATGGCGGGAAGATGAAGACA 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        194 CACACAACAGAGGCAGTTCCAGATTTCAACTGCTCATCAGATGGCGGGAAGATGAAGACA 135

Qy        183 GATGGTGCAGCCACAGTTCGTTTGATCTCCAGCTCGAGCC 222
              ||||||||||||||||||||||||||||||| || |  ||
Db        134 GATGGTGCAGCCACAGTTCGTTTGATCTCCACCTTGGTCC 95


RESULT 12
US-08-157-101A-8/c
; Sequence 8, Application US/08157101A
; Patent No. 5808032
;  GENERAL INFORMATION:
;    APPLICANT:  KURIHARA, TATSUYA
;    APPLICANT:  MATSUKURA, SHIGEKAZU
;    APPLICANT:  TSURUOKA, NOBUO
;    APPLICANT:  ARIMA, KENJI
;    APPLICANT:  NISHIHARA, TATSURO
;    TITLE OF INVENTION:  ANTI-HBs ANTIBODY GENES AND EXPRESSION
;    TITLE OF INVENTION:  PLASMIDS THEREFOR
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PILLSBURY, MADISON & SUTRO
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/157,101A
;      FILING DATE:  05-APR-1994
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  TITUS, MARLANA K
;      REGISTRATION NUMBER:  35843
;      REFERENCE/DOCKET NUMBER:  9437/204199
```

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3711
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUCH
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  642 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-157-101A-8

  Query Match             36.4%;  Score 212;  DB 1;  Length 642;
  Best Local Similarity   97.7%;  Pred. No. 2.9e-55;
  Matches  215;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          3 CTGTAGGTGCTGTCCTTGCTGTCCTGCTCTGTGACACTCTCCTGGGAGTCACCCGATTGG 62
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        521 CTGTAGGTGCTGTCCTTGCTGTCCTGCTCTGTGACACTCTCCTGGGAGTTACCCGATTGG 462

Qy         63 AGGGCGTTATCCACCTTCCACTGTACTTTGGCCTCTCTGGGATAGAAGTTATTCAGCAGG 122
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 AGGGCGTTATCCACCTTCCACTGTACTTTGGCCTCTCTGGGATAGAAGTTATTCAGCAGG 402

Qy        123 CACACAACAGAGGCAGTTCCAGATTTCAACTGCTCATCAGATGGCGGGAAGATGAAGACA 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 CACACAACAGAGGCAGTTCCAGATTTCAACTGCTCATCAGATGGCGGGAAGATGAAGACA 342

Qy        183 GATGGTGCAGCCACAGTTCGTTTGATCTCCAGCTCGAGCC 222
              |||||||||||||||||||||||||||||| || |  ||
Db        341 GATGGTGCAGCCACAGTTCGTTTGATCTCCACCTTGGTCC 302


RESULT 13
US-08-157-101A-4/c
; Sequence 4, Application US/08157101A
; Patent No. 5808032
;  GENERAL INFORMATION:
;    APPLICANT:  KURIHARA, TATSUYA
;    APPLICANT:  MATSUKURA, SHIGEKAZU
;    APPLICANT:  TSURUOKA, NOBUO
;    APPLICANT:  ARIMA, KENJI
;    APPLICANT:  NISHIHARA, TATSURO
;    TITLE OF INVENTION:  ANTI-HBs ANTIBODY GENES AND EXPRESSION
;    TITLE OF INVENTION:  PLASMIDS THEREFOR
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PILLSBURY, MADISON & SUTRO
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/157,101A
;      FILING DATE:  05-APR-1994
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  TITUS, MARLANA K
;      REGISTRATION NUMBER:  35843
;      REFERENCE/DOCKET NUMBER:  9437/204199
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3711
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUCH
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1066 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-157-101A-4

  Query Match             36.4%;  Score 212;  DB 1;  Length 1066;
  Best Local Similarity   97.7%;  Pred. No. 3.8e-55;
  Matches  215;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          3 CTGTAGGTGCTGTCCTTGCTGTCCTGCTCTGTGACACTCTCCTGGGAGTCACCCGATTGG 62
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        619 CTGTAGGTGCTGTCCTTGCTGTCCTGCTCTGTGACACTCTCCTGGGAGTTACCCGATTGG 560

Qy         63 AGGGCGTTATCCACCTTCCACTGTACTTTGGCCTCTCTGGGATAGAAGTTATTCAGCAGG 122
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        559 AGGGCGTTATCCACCTTCCACTGTACTTTGGCCTCTCTGGGATAGAAGTTATTCAGCAGG 500

Qy        123 CACACAACAGAGGCAGTTCCAGATTTCAACTGCTCATCAGATGGCGGGAAGATGAAGACA 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        499 CACACAACAGAGGCAGTTCCAGATTTCAACTGCTCATCAGATGGCGGGAAGATGAAGACA 440

Qy        183 GATGGTGCAGCCACAGTTCGTTTGATCTCCAGCTCGAGCC 222
              |||||||||||||||||||||||||||||| || |  ||
Db        439 GATGGTGCAGCCACAGTTCGTTTGATCTCCACCTTGGTCC 400


RESULT 14
US-09-171-945-51/c
; Sequence 51, Application US/09171945
; Patent No. 6277599
; GENERAL INFORMATION:
;  APPLICANT:  Emery, Stephen
;  APPLICANT:  Copley, Clive Graham
;  APPLICANT:  Edge, Michael Derek
;  TITLE OF INVENTION:  Monoclonal Antibody to CEA, Conjugates Comprising Said
;  TITLE OF INVENTION:  Antibody, and Their Therapeutic Use in an Adept System
;  FILE REFERENCE:  Monoclonal Antibody to CEA
;  CURRENT APPLICATION NUMBER: US/09/171,945
;  CURRENT FILING DATE:  1998-10-29
;  PRIOR APPLICATION NUMBER:  GB9703103.3
;  PRIOR FILING DATE:  1997-02-14
;  PRIOR APPLICATION NUMBER:  GB9609405.7
;  PRIOR FILING DATE:  1996-05-04
;  PRIOR APPLICATION NUMBER:  PCT/GB97/01165
;  PRIOR FILING DATE:  1997-04-29
;  NUMBER OF SEQ ID NOS:  131
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 51
;   LENGTH: 705
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: humanized
US-09-171-945-51

  Query Match             36.2%;  Score 211.2;  DB 4;  Length 705;
  Best Local Similarity   96.4%;  Pred. No. 5.4e-55;
  Matches  216;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          3 CTGTAGGTGCTGTCCTTGCTGTCCTGCTCTGTGACACTCTCCTGGGAGTCACCCGATTGG 62
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        584 CTGTAGGTGCTGTCCTTGCTGTCCTGCTCTGTGACACTCTCCTGGGAGTTACCCGATTGG 525

Qy         63 AGGGCGTTATCCACCTTCCACTGTACTTTGGCCTCTCTGGGATAGAAGTTATTCAGCAGG 122
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        524 AGGGCGTTATCCACCTTCCACTGTACTTTGGCCTCTCTGGGATAGAAGTTATTCAGCAGG 465

Qy        123 CACACAACAGAGGCAGTTCCAGATTTCAACTGCTCATCAGATGGCGGGAAGATGAAGACA 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        464 CACACAACAGAGGCAGTTCCAGATTTCAACTGCTCATCAGATGGCGGGAAGATGAAGACA 405
```

```
Qy        183 GATGGTGCAGCCACAGTTCGTTTGATCTCCAGCTCGAGCCGCTG 226
              ||||||||||||||||| |||||||||| |||| |  ||  ||
Db        404 GATGGTGCAGCCACAGTCCGTTTGATCTCGAGCTTGGTCCCTTG 361


RESULT 15
US-09-171-945-96/c
; Sequence 96, Application US/09171945
; Patent No. 6277599
; GENERAL INFORMATION:
;  APPLICANT:  Emery, Stephen
;  APPLICANT:  Copley, Clive Graham
;  APPLICANT:  Edge, Michael Derek
;  TITLE OF INVENTION:  Monoclonal Antibody to CEA, Conjugates Comprising Said
;  TITLE OF INVENTION:  Antibody, and Their Therapeutic Use in an Adept System
;  FILE REFERENCE:  Monoclonal Antibody to CEA
;  CURRENT APPLICATION NUMBER: US/09/171,945
;  CURRENT FILING DATE:  1998-10-29
;  PRIOR APPLICATION NUMBER:  GB9703103.3
;  PRIOR FILING DATE:  1997-02-14
;  PRIOR APPLICATION NUMBER:  GB9609405.7
;  PRIOR FILING DATE:  1996-05-04
;  PRIOR APPLICATION NUMBER:  PCT/GB97/01165
;  PRIOR FILING DATE:  1997-04-29
;  NUMBER OF SEQ ID NOS:  131
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 96
;   LENGTH: 705
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: humanized
US-09-171-945-96

  Query Match             36.2%;  Score 211.2;  DB 4;  Length 705;
  Best Local Similarity   96.4%;  Pred. No. 5.4e-55;
  Matches  216;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          3 CTGTAGGTGCTGTCCTTGCTGTCCTGCTCTGTGACACTCTCCTGGGAGTCACCCGATTGG 62
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        584 CTGTAGGTGCTGTCCTTGCTGTCCTGCTCTGTGACACTCTCCTGGGAGTTACCCGATTGG 525

Qy         63 AGGGCGTTATCCACCTTCCACTGTACTTTGGCCTCTCTGGGATAGAAGTTATTCAGCAGG 122
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        524 AGGGCGTTATCCACCTTCCACTGTACTTTGGCCTCTCTGGGATAGAAGTTATTCAGCAGG 465

Qy        123 CACACAACAGAGGCAGTTCCAGATTTCAACTGCTCATCAGATGGCGGGAAGATGAAGACA 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        464 CACACAACAGAGGCAGTTCCAGATTTCAACTGCTCATCAGATGGCGGGAAGATGAAGACA 405

Qy        183 GATGGTGCAGCCACAGTTCGTTTGATCTCCAGCTCGAGCCGCTG 226
              ||||||||||||||||| |||||||||| |||| |  ||  ||
Db        404 GATGGTGCAGCCACAGTCCGTTTGATCTCGAGCTTGGTCCCTTG 361


Search completed: July 16, 2003, 14:58:51
Job time : 23.7443 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 16, 2003, 13:28:07 ; Search time 76.7139 Seconds
(without alignments)
12133.023 Million cell updates/sec

Title: US-09-981-353-114
Perfect score: 583
Sequence: 1 ccctgtaggtgctgtccttg.....tatgtgaatgataatagtca 583

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1114047 seqs, 798260406 residues

Total number of hits satisfying chosen parameters: 2228094

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
6: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	583	100.0	583	9	US-09-981-353-114	Sequence 114, App
	2	583	100.0	583	9	US-10-084-817-179	Sequence 179, App
	3	583	100.0	583	12	US-10-044-090-584	Sequence 584, App
	4	277.8	47.7	448	9	US-10-125-237-66	Sequence 66, Appl
	5	277.8	47.7	448	9	US-10-105-891-66	Sequence 66, Appl
	6	277.4	47.6	415	10	US-09-880-107-3794	Sequence 3794, Ap
	7	275.8	47.3	473	9	US-09-919-039-244	Sequence 244, App
c	8	245.2	42.1	409	9	US-09-954-531-1253	Sequence 1253, Ap
c	9	245.2	42.1	409	10	US-09-880-107-2702	Sequence 2702, Ap
	10	242.6	41.6	459	9	US-09-918-995-30613	Sequence 30613, A
	11	238.2	40.9	457	9	US-09-918-995-32599	Sequence 32599, A
	12	234.8	40.3	409	12	US-10-044-090-524	Sequence 524, App
	13	234.4	40.2	523	12	US-10-044-090-835	Sequence 835, App
	14	232.8	39.9	453	10	US-09-925-300-821	Sequence 821, App
	15	228.2	39.1	409	9	US-09-918-995-8793	Sequence 8793, Ap
	16	227.6	39.0	442	9	US-09-981-353-119	Sequence 119, App
	17	227.6	39.0	442	9	US-09-919-039-271	Sequence 271, App
	18	227.6	39.0	442	12	US-10-044-090-525	Sequence 525, App
	19	227.6	39.0	525	9	US-10-125-237-16	Sequence 16, Appl
	20	227.6	39.0	525	9	US-10-105-891-16	Sequence 16, Appl
	21	227.6	39.0	682	10	US-09-925-300-490	Sequence 490, App
	22	226.2	38.8	558	9	US-09-919-039-194	Sequence 194, App
	23	226.2	38.8	558	9	US-10-084-817-119	Sequence 119, App
	24	223.6	38.4	424	9	US-10-119-926-71	Sequence 71, Appl
	25	219.8	37.7	367	10	US-09-880-107-3760	Sequence 3760, Ap
	26	219.4	37.6	493	12	US-10-044-090-847	Sequence 847, App
	27	219.2	37.6	466	9	US-09-918-995-31299	Sequence 31299, A
	28	217.4	37.3	497	9	US-09-918-995-32025	Sequence 32025, A
	29	216.4	37.1	500	10	US-09-925-297-454	Sequence 454, App
c	30	216	37.0	657	9	US-09-972-656-91	Sequence 91, Appl
c	31	216	37.0	657	9	US-09-972-656-103	Sequence 103, App
c	32	216	37.0	657	9	US-09-972-656-105	Sequence 105, App
c	33	216	37.0	678	9	US-09-453-234-73	Sequence 73, Appl
c	34	216	37.0	678	9	US-09-453-234-79	Sequence 79, Appl
c	35	216	37.0	678	9	US-09-453-234-85	Sequence 85, Appl
c	36	216	37.0	948	10	US-09-859-053-33	Sequence 33, Appl
c	37	216	37.0	970	10	US-09-859-053-37	Sequence 37, Appl
c	38	216	37.0	1202	9	US-10-158-646-57	Sequence 57, Appl
	39	216	37.0	2272	9	US-10-158-646-63	Sequence 63, Appl
c	40	215.4	36.9	829	9	US-10-158-646-4	Sequence 4, Appli
c	41	214.6	36.8	1045	9	US-10-198-846-13629	Sequence 13629, A
c	42	214.4	36.8	678	9	US-09-453-234-49	Sequence 49, Appl
c	43	214.4	36.8	1033	10	US-09-799-514-2	Sequence 2, Appli
c	44	214.4	36.8	1230	9	US-10-158-646-59	Sequence 59, Appl
c	45	214.4	36.8	2272	9	US-10-158-646-63	Sequence 63, Appl

ALIGNMENTS

RESULT 1
US-09-981-353-114
; Sequence 114, Application US/09981353
; Patent No. US20020160382A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy W.
; APPLICANT: Jones, David A.
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
; FILE REFERENCE: PA-0038 US
; CURRENT APPLICATION NUMBER: US/09/981,353
; CURRENT FILING DATE: 2001-10-11
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 114
; LENGTH: 583
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020160382A1 2685676CB1
US-09-981-353-114

Query Match 100.0%; Score 583; DB 9; Length 583;
Best Local Similarity 100.0%; Pred. No. 7.5e-179;
Matches 583; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 CCCTGTAGGTGCTGTCCTTGCTGTCCTGCTCTGTGACACTCTCCTGGGAGTCACCCGATT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCCTGTAGGTGCTGTCCTTGCTGTCCTGCTCTGTGACACTCTCCTGGGAGTCACCCGATT 60

Qy         61 GGAGGGCGTTATCCACCTTCCACTGTACTTTGGCCTCTCTGGGATAGAAGTTATTCAGCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGAGGGCGTTATCCACCTTCCACTGTACTTTGGCCTCTCTGGGATAGAAGTTATTCAGCA 120

Qy        121 GGCACACAACAGAGGCAGTTCCAGATTTCAACTGCTCATCAGATGGCGGGAAGATGAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGCACACAACAGAGGCAGTTCCAGATTTCAACTGCTCATCAGATGGCGGGAAGATGAAGA 180

Qy        181 CAGATGGTGCAGCCACAGTTCGTTTGATCTCCAGCTCGAGCCGCTGCGTGTTTTCCTCTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGGTGCAGCCACAGTTCGTTTGATCTCCAGCTCGAGCCGCTGCGTGTTTTCCTCTT 240
```

```
Qy        241 GATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCCCAACTGCTCCTGCTCGCCTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCCCAACTGCTCCTGCTCGCCTG 300

Qy        301 TTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGAGTGCAAATGCACCTCCTGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGAGTGCAAATGCACCTCCTGCA 360

Qy        361 AGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAAGTGTGCCCAGGGCTGCATCT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAAGTGTGCCCAGGGCTGCATCT 420

Qy        421 GCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATGTCCGGACAGCCCTGCTCGAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATGTCCGGACAGCCCTGCTCGAA 480

Qy        481 GATATAGAAAGAGTGACCTGCACAAACTTGGAATTTTTTTTCCATACAACCCTGCCCCAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GATATAGAAAGAGTGACCTGCACAAACTTGGAATTTTTTTTCCATACAACCCTGCCCCAT 540

Qy        541 CTACTGTATTTTTTTTAATGAAATATGTGAATGATAATAGTCA 583
              |||||||||||||||||||||||||||||||||||||||||||
Db        541 CTACTGTATTTTTTTTAATGAAATATGTGAATGATAATAGTCA 583


RESULT 2
US-10-084-817-179
; Sequence 179, Application US/10084817
; Publication No. US20030119009A1
; GENERAL INFORMATION:
;  APPLICANT: Susan Stuart
;  APPLICANT:  Jed G. Nuchtern
;  APPLICANT:  Sharon E. Plon
;  APPLICANT:  Jason M. Shohet
;  TITLE OF INVENTION: GENES REGULATED BY MYCN ACTIVATION
;  FILE REFERENCE: PA-0046 US
;  CURRENT APPLICATION NUMBER: US/10/084,817
;  CURRENT FILING DATE:  2002-02-25
;  PRIOR APPLICATION NUMBER: 60/270,784
;  PRIOR FILING DATE: 2001-02-23
;  NUMBER OF SEQ ID NOS: 365
;  SOFTWARE: PERL Program
; SEQ ID NO 179
;   LENGTH: 583
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20030119009A1 2685676CB1
US-10-084-817-179

  Query Match            100.0%;  Score 583;  DB 9;  Length 583;
  Best Local Similarity  100.0%;  Pred. No. 7.5e-179;
  Matches  583;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCCTGTAGGTGCTGTCCTTGCTGTCCTGCTCTGTGACACTCTCCTGGGAGTCACCCGATT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCCTGTAGGTGCTGTCCTTGCTGTCCTGCTCTGTGACACTCTCCTGGGAGTCACCCGATT 60

Qy         61 GGAGGGCGTTATCCACCTTCCACTGTACTTTGGCCTCTCTGGGATAGAAGTTATTCAGCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGAGGGCGTTATCCACCTTCCACTGTACTTTGGCCTCTCTGGGATAGAAGTTATTCAGCA 120

Qy        121 GGCACACAACAGAGGCAGTTCCAGATTTCAACTGCTCATCAGATGGCGGGAAGATGAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGCACACAACAGAGGCAGTTCCAGATTTCAACTGCTCATCAGATGGCGGGAAGATGAAGA 180

Qy        181 CAGATGGTGCAGCCACAGTTCGTTTGATCTCCAGCTCGAGCCGCTGCGTGTTTTCCTCTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGGTGCAGCCACAGTTCGTTTGATCTCCAGCTCGAGCCGCTGCGTGTTTTCCTCTT 240

Qy        241 GATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCCCAACTGCTCCTGCTCGCCTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCCCAACTGCTCCTGCTCGCCTG 300

Qy        301 TTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGAGTGCAAATGCACCTCCTGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGAGTGCAAATGCACCTCCTGCA 360

Qy        361 AGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAAGTGTGCCCAGGGCTGCATCT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAAGTGTGCCCAGGGCTGCATCT 420

Qy        421 GCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATGTCCGGACAGCCCTGCTCGAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATGTCCGGACAGCCCTGCTCGAA 480

Qy        481 GATATAGAAAGAGTGACCTGCACAAACTTGGAATTTTTTTTCCATACAACCCTGCCCCAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GATATAGAAAGAGTGACCTGCACAAACTTGGAATTTTTTTTCCATACAACCCTGCCCCAT 540

Qy        541 CTACTGTATTTTTTTTAATGAAATATGTGAATGATAATAGTCA 583
              |||||||||||||||||||||||||||||||||||||||||||
Db        541 CTACTGTATTTTTTTTAATGAAATATGTGAATGATAATAGTCA 583


RESULT 3
US-10-044-090-584
; Sequence 584, Application US/10044090
; Patent No. US20020137081A1
; GENERAL INFORMATION:
;  APPLICANT: Olga Bandman
;  TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
;  FILE REFERENCE: PA-0028 US
;  CURRENT APPLICATION NUMBER: US/10/044,090
;  CURRENT FILING DATE:  2002-01-09
;  NUMBER OF SEQ ID NOS: 850
;  SOFTWARE: PERL Program
; SEQ ID NO 584
;   LENGTH: 583
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20020137081A1 2685676CB1
US-10-044-090-584

  Query Match            100.0%;  Score 583;  DB 12;  Length 583;
  Best Local Similarity  100.0%;  Pred. No. 7.5e-179;
  Matches  583;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCCTGTAGGTGCTGTCCTTGCTGTCCTGCTCTGTGACACTCTCCTGGGAGTCACCCGATT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCCTGTAGGTGCTGTCCTTGCTGTCCTGCTCTGTGACACTCTCCTGGGAGTCACCCGATT 60

Qy         61 GGAGGGCGTTATCCACCTTCCACTGTACTTTGGCCTCTCTGGGATAGAAGTTATTCAGCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGAGGGCGTTATCCACCTTCCACTGTACTTTGGCCTCTCTGGGATAGAAGTTATTCAGCA 120

Qy        121 GGCACACAACAGAGGCAGTTCCAGATTTCAACTGCTCATCAGATGGCGGGAAGATGAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGCACACAACAGAGGCAGTTCCAGATTTCAACTGCTCATCAGATGGCGGGAAGATGAAGA 180

Qy        181 CAGATGGTGCAGCCACAGTTCGTTTGATCTCCAGCTCGAGCCGCTGCGTGTTTTCCTCTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAGATGGTGCAGCCACAGTTCGTTTGATCTCCAGCTCGAGCCGCTGCGTGTTTTCCTCTT 240

Qy        241 GATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCCCAACTGCTCCTGCTCGCCTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCCCAACTGCTCCTGCTCGCCTG 300

Qy        301 TTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGAGTGCAAATGCACCTCCTGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGAGTGCAAATGCACCTCCTGCA 360

Qy        361 AGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAAGTGTGCCCAGGGCTGCATCT 420
```

```
Db        361 AGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAAGTGTGCCCAGGGCTGCATCT 420

Qy        421 GCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATGTCCGGACAGCCCTGCTCGAA 480

Db        421 GCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATGTCCGGACAGCCCTGCTCGAA 480

Qy        481 GATATAGAAAGAGTGACCTGCACAAACTTGGAATTTTTTTTCCATACAACCCTGCCCCAT 540

Db        481 GATATAGAAAGAGTGACCTGCACAAACTTGGAATTTTTTTTCCATACAACCCTGCCCCAT 540

Qy        541 CTACTGTATTTTTTTTAATGAAATATGTGAATGATAATAGTCA 583

Db        541 CTACTGTATTTTTTTTAATGAAATATGTGAATGATAATAGTCA 583
```

RESULT 4
US-10-125-237-66
; Sequence 66, Application US/10125237
; Publication No. US20030022329A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Zhou, Ping
; APPLICANT: Asundi, Vinod
; APPLICANT: Ren, Feiyan
; APPLICANT: Zhao, Qing A.
; APPLICANT: Xue, Aidong J.
; APPLICANT: Zhang, Jie
; APPLICANT: Wehrman, Tom
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20030022329A1el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 791CIP2ADIV
; CURRENT APPLICATION NUMBER: US/10/125,237
; CURRENT FILING DATE: 2002-04-17
; PRIOR APPLICATION NUMBER: 09/668,317
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: 09/552,929
; PRIOR FILING DATE: 2000-04-18
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: pt_FL_genes Version 2.0
; SEQ ID NO 66
; LENGTH: 448
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (106)..(291)
US-10-125-237-66

Query Match 47.7%; Score 277.8; DB 9; Length 448;
Best Local Similarity 88.6%; Pred. No. 8.1e-80;
Matches 327; Conservative 0; Mismatches 32; Indels 10; Gaps 2;

```
Qy        221 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 280

Db         54 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 113

Qy        281 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 340

Db        114 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 173

Qy        341 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAA 400

Db        174 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAA 233

Qy        401 GTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATG 460

Db        234 GTGTGCCCAGGGCTGCATCTGCAAAGGGACGTCAGACAAGTGCAGCTGCTGTGCCTGATG 293

Qy        461 TCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGA-------- 512

Db        294 CCAGGACAGCTGTGCTCTCAGATGTAAATAGAGCAACCTATATAAACCTGGATTTTTTTT 353

Qy        513 ATTTTTTTTCCATACAACCCTGCCCCATCTACT--GTATTTTTTTTAATGAAATATGTGA 570

Db        354 TTTTTTTTTTTGTACAACCCTGACCCGTTTGCTACATCTTTTTTTCTATGAAATATGTGA 413

Qy        571 ATGATAATA 579

Db        414 ATGGCAATA 422
```

RESULT 5
US-10-105-891-66
; Sequence 66, Application US/10105891
; Publication No. US20030073099A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Zhou, Ping
; APPLICANT: Asundi, Vinod
; APPLICANT: Ren, Feiyan
; APPLICANT: Zhao, Qing A.
; APPLICANT: Xue, Aidong J.
; APPLICANT: Zhang, Jie
; APPLICANT: Wehrman, Tom
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20030073099A1el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 791CIP2A
; CURRENT APPLICATION NUMBER: US/10/105,891
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER: 09/668,317
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: 09/552,929
; PRIOR FILING DATE: 2000-04-18
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: pt_FL_genes Version 2.0
; SEQ ID NO 66
; LENGTH: 448
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (106)..(291)
US-10-105-891-66

Query Match 47.7%; Score 277.8; DB 9; Length 448;
Best Local Similarity 88.6%; Pred. No. 8.1e-80;
Matches 327; Conservative 0; Mismatches 32; Indels 10; Gaps 2;

```
Qy        221 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 280

Db         54 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 113

Qy        281 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 340

Db        114 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 173

Qy        341 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAA 400

Db        174 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAA 233

Qy        401 GTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATG 460

Db        234 GTGTGCCCAGGGCTGCATCTGCAAAGGGACGTCAGACAAGTGCAGCTGCTGTGCCTGATG 293

Qy        461 TCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGA-------- 512

Db        294 CCAGGACAGCTGTGCTCTCAGATGTAAATAGAGCAACCTATATAAACCTGGATTTTTTTT 353

Qy        513 ATTTTTTTTCCATACAACCCTGCCCCATCTACT--GTATTTTTTTTAATGAAATATGTGA 570
```

```
              |||||||||   |||||||||| ||| | | ||   | |||||||  ||||||||||||
Db        354 TTTTTTTTTTTGTACAACCCTGACCCGTTTGCTACATCTTTTTTTCTATGAAATATGTGA 413

Qy        571 ATGATAATA 579
              |||  ||||
Db        414 ATGGCAATA 422


RESULT 6
US-09-880-107-3794
; Sequence 3794, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
;  APPLICANT: Horne, Darci T.
;  APPLICANT:  Vockley, Joseph G.
;  APPLICANT:  Scherf, Uwe
;  APPLICANT:  Gene Logic, Inc.
;  TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
;  FILE REFERENCE: 44921-5028-WO
;  CURRENT APPLICATION NUMBER: US/09/880,107
;  CURRENT FILING DATE:  2001-06-14
;  PRIOR APPLICATION NUMBER: US 60/211,379
;  PRIOR FILING DATE: 2000-06-14
;  PRIOR APPLICATION NUMBER: US 60/237,054
;  PRIOR FILING DATE: 2000-10-02
;  NUMBER OF SEQ ID NOS: 3950
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3794
;   LENGTH: 415
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. US20020142981A1 X76717
US-09-880-107-3794

  Query Match             47.6%;  Score 277.4;  DB 10;  Length 415;
  Best Local Similarity   88.3%;  Pred. No. 1e-79;
  Matches  326;  Conservative    1;  Mismatches   32;  Indels   10;  Gaps    2;

Qy        221 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 60

Qy        281 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 120

Qy        341 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAA 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:||
Db        121 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCMAA 180

Qy        401 GTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATG 460
              |||||||||||||||||||||||||| | ||||| ||||||||||||||||||||||||
Db        181 GTGTGCCCAGGGCTGCATCTGCAAAGGGACGTCAGACAAGTGCAGCTGCTGTGCCTGATG 240

Qy        461 TCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGA-------- 512
              | |||||||  |||||  |||| || | ||||  ||||  | |||| ||||
Db        241 CCAGGACAGCTGTGCTCTCAGATGTAAATAGAGCAACCTATATAAACCTGGATTTTTTTT 300

Qy        513 ATTTTTTTTCCATACAACCCTGCCCCATCTACT--GTATTTTTTTTAATGAAATATGTGA 570
              ||||||||   |||||||||| ||| | | ||   | |||||||  ||||||||||||
Db        301 TTTTTTTTTTTGTACAACCCTGACCCGTTTGCTACATCTTTTTTTCTATGAAATATGTGA 360

Qy        571 ATGATAATA 579
              |||  ||||
Db        361 ATGGCAATA 369


RESULT 7
US-09-919-039-244
; Sequence 244, Application US/09919039
; Publication No. US20030108871A1 .
; GENERAL INFORMATION:
;  APPLICANT: Kaser, Matthew R.
;  TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
;  FILE REFERENCE: PA-0035 US
;  CURRENT APPLICATION NUMBER: US/09/919,039
;  CURRENT FILING DATE:  2002-09-09
;  PRIOR APPLICATION NUMBER: 60/222,113
;  PRIOR FILING DATE: 2000-07-28
;  NUMBER OF SEQ ID NOS: 401
;  SOFTWARE: PERL Program
; SEQ ID NO 244
;   LENGTH: 473
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20030108871A1 167772CB1
US-09-919-039-244

  Query Match             47.3%;  Score 275.8;  DB 9;  Length 473;
  Best Local Similarity   88.1%;  Pred. No. 3.7e-79;
  Matches  327;  Conservative    0;  Mismatches   32;  Indels   12;  Gaps    2;

Qy        221 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         56 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 115

Qy        281 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        116 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 175

Qy        341 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAA 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        176 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAA 235

Qy        401 GTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATG 460
              |||||||||||||||||||||||||| | ||||| ||||||||||||||||||||||||
Db        236 GTGTGCCCAGGGCTGCATCTGCAAAGGGACGTCAGACAAGTGCAGCTGCTGTGCCTGATG 295

Qy        461 TCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGA-------- 512
              | |||||||  |||||  |||| || | ||||  ||||  | |||| ||||
Db        296 CCAGGACAGCTGTGCTCTCAGATGTAAATAGAGCAACCTATATAAACCTGGATTTTTTTT 355

Qy        513 --ATTTTTTTTCCATACAACCCTGCCCCATCTACT--GTATTTTTTTTAATGAAATATGT 568
                |||||||   |||||||||| ||| | | ||   | |||||||  ||||||||||
Db        356 TTTTTTTTTTTTGTACAACCCTGACCCGTTTGCTACATCTTTTTTTCTATGAAATATGT 415

Qy        569 GAATGATAATA 579
              |||||  ||||
Db        416 GAATGGCAATA 426


RESULT 8
US-09-954-531-1253/c
; Sequence 1253, Application US/09954531
; Patent No. US20020165180A1
; GENERAL INFORMATION:
;  APPLICANT: Weaver, Zoe
;  TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
;  TITLE OF INVENTION:  Gene Sets
;  FILE REFERENCE: 689290-77
;  CURRENT APPLICATION NUMBER: US/09/954,531
;  CURRENT FILING DATE:  2002-05-02
;  PRIOR APPLICATION NUMBER: US/60/233,133
;  PRIOR FILING DATE: 2000-09-18
;  PRIOR APPLICATION NUMBER: US/60/234,009
;  PRIOR FILING DATE: 2000-09-20
;  PRIOR APPLICATION NUMBER: US/60/234,034
;  PRIOR FILING DATE: 2000-09-20
;  PRIOR APPLICATION NUMBER: US/60/234,509
;  PRIOR FILING DATE: 2000-09-22
;  PRIOR APPLICATION NUMBER: US/60/234,567
;  PRIOR FILING DATE: 2000-09-22
;  NUMBER OF SEQ ID NOS: 1392
```

```
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1253
;   LENGTH: 409
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(409)
;   OTHER INFORMATION: n=a,t,g or c
US-09-954-531-1253

  Query Match            42.1%;  Score 245.2;  DB 9;  Length 409;
  Best Local Similarity  88.0%;  Pred. No. 3e-69;
  Matches  324;  Conservative    0;  Mismatches   35;  Indels    9;  Gaps    5;

Qy        221 CCGCTGCGTGTTTTCCTCTTG-ATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACC 279
              |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db        382 CCGCTGCGTGTTTTCCTCTTGAATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACC 323

Qy        280 CCAACTGCTCCTGC-TCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAA 338
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        322 CCAACTGCTCCTGCNTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAA 263

Qy        339 GAGTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCC-TGCTGCCCTGTGGGCTGTGC 397
              |||||||||||||||||||||||||||||||||||||| |||   ||||||||||||||
Db        262 GAGTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCNTGCNTNCCTGTGGGCTGTGC 203

Qy        398 CAAGTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTG 457
              ||||||||||||||||||||||||||||| | ||||| ||||||||||||||||||||
Db        202 CAAGTGTGCCCAGGGCTGCATCTGCAAAGGGACGTCAGACAAGTGCAGCTGCTGTGCCTG 143

Qy        458 ATGTCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGA----A 513
              ||| | |||||||  |||||  |||| || | ||||  ||||  | |||| ||||
Db        142 ATGCCAGGACAGCTGTGCTCTCAGATGTAAATAGAGCAACCTATATAAACCTGGATTTTT 83

Qy        514 TTTTTTTTCCATACAACCCTGCCCCATCTACT--GTATTTTTTTTAATGAAATATGTGAA 571
              ||||||||   |||||||||| ||| | | ||   | |||||||  ||||||||||||||
Db         82 TTTTTTTTTTGTACAACCCTGACCCGTTTGCTACATCTTTTTTTCTATGAAATATGTGAA 23

Qy        572 TGATAATA 579
              ||  ||||
Db         22 TGGCAATA 15


RESULT 9
US-09-880-107-2702/c
; Sequence 2702, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
;  APPLICANT: Horne, Darci T.
;  APPLICANT:  Vockley, Joseph G.
;  APPLICANT:  Scherf, Uwe
;  APPLICANT:  Gene Logic, Inc.
;  TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
;  FILE REFERENCE: 44921-5028-WO
;  CURRENT APPLICATION NUMBER: US/09/880,107
;  CURRENT FILING DATE:  2001-06-14
;  PRIOR APPLICATION NUMBER: US 60/211,379
;  PRIOR FILING DATE: 2000-06-14
;  PRIOR APPLICATION NUMBER: US 60/237,054
;  PRIOR FILING DATE: 2000-10-02
;  NUMBER OF SEQ ID NOS: 3950
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2702
;   LENGTH: 409
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. US20020142981A1 N80129
;   NAME/KEY: unsure
;   LOCATION: (1)..(409)
;   OTHER INFORMATION: n = a or c or g or t
US-09-880-107-2702

  Query Match            42.1%;  Score 245.2;  DB 10;  Length 409;
  Best Local Similarity  88.0%;  Pred. No. 3e-69;
  Matches  324;  Conservative    0;  Mismatches   35;  Indels    9;  Gaps    5;

Qy        221 CCGCTGCGTGTTTTCCTCTTG-ATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACC 279
              |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db        382 CCGCTGCGTGTTTTCCTCTTGAATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACC 323

Qy        280 CCAACTGCTCCTGC-TCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAA 338
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        322 CCAACTGCTCCTGCNTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAA 263

Qy        339 GAGTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCC-TGCTGCCCTGTGGGCTGTGC 397
              |||||||||||||||||||||||||||||||||||||| |||   ||||||||||||||
Db        262 GAGTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCNTGCNTNCCTGTGGGCTGTGC 203

Qy        398 CAAGTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTG 457
              ||||||||||||||||||||||||||||| | ||||| ||||||||||||||||||||
Db        202 CAAGTGTGCCCAGGGCTGCATCTGCAAAGGGACGTCAGACAAGTGCAGCTGCTGTGCCTG 143

Qy        458 ATGTCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGA----A 513
              ||| | |||||||  |||||  |||| || | ||||  ||||  | |||| ||||
Db        142 ATGCCAGGACAGCTGTGCTCTCAGATGTAAATAGAGCAACCTATATAAACCTGGATTTTT 83

Qy        514 TTTTTTTTCCATACAACCCTGCCCCATCTACT--GTATTTTTTTTAATGAAATATGTGAA 571
              ||||||||   |||||||||| ||| | | ||   | |||||||  ||||||||||||||
Db         82 TTTTTTTTTTGTACAACCCTGACCCGTTTGCTACATCTTTTTTTCTATGAAATATGTGAA 23

Qy        572 TGATAATA 579
              ||  ||||
Db         22 TGGCAATA 15


RESULT 10
US-09-918-995-30613
; Sequence 30613, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-756
;  CURRENT APPLICATION NUMBER: US/09/918,995
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US/09/235,076
;  PRIOR FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 38054
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30613
;   LENGTH: 459
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(459)
;   OTHER INFORMATION: n = A,T,C or G
US-09-918-995-30613

  Query Match            41.6%;  Score 242.6;  DB 9;  Length 459;
  Best Local Similarity  79.6%;  Pred. No. 2.2e-68;
  Matches  313;  Conservative    0;  Mismatches   74;  Indels    6;  Gaps    2;

Qy        191 AGCCACAGTTCGTTTGATCTCCAGCTCGAGCCGCTGCGTGTTTTCCTCTTGATCGGGAAC 250
              | ||| |||  || || |      |  |  || |||| | || || ||| | | ||||||
Db         32 ATCCATAGTCAGTGTGGTGGAATTCGTGTTCCACTGCCTCTTCTCTTCTCGCTTGGGAAC 91

Qy        251 TCCTGCTTCTCCTTGCCTCGAAATGGACCCCAACTGCTCCTGCTCGCCTGTTGGCTCCTG 310
              ||| |  || ||| | || | ||||||||||||||||||||||  | ||| ||||||||
Db         92 TCCAGTCTCACCTCGGCTTGCAATGGACCCCAACTGCTCCTGCGAGGCTGGTGGCTCCTG 151

Qy        311 TGCCTGTGCCGGCTCCTGCAAATGCAAAGAGTGCAAATGCACCTCCTGCAAGAAGAGCTG 370
              ||||| |||||||||||||| |||||| ||||||||||||||||||||||||||||||
```

```
Db        152 CGCCTGCGCCGGCTCCTGCAAGTGCAAAAAGTGCAAATGCACCTCCTGCAAGAAGAGCTG 211

Qy        371 CTGCTCCTGCTGCCCTGTGGGCTGTGCCAAGTGTGCCCAGGGCTGCATCTGCAAAGGGGC 430
              |||||||| |||||  ||||||||||||||||||||||||||||||||||||||||||||
Db        212 CTGCTCCTGTTGCCCCCTGGGCTGTGCCAAGTGTGCCCAGGGCTGCATCTGCAAAGGGGC 271

Qy        431 ATCAGAGAAGTGCAGCTGCTGTGCCTGATGTCCGGACAGCCCTGCTCGAAGATATAGAAA 490
              |||||||||||||||||||||||||||||||| ||||||||||||   ||||  | | |
Db        272 GTCAGAGAAGTGCAGCTGCTGTGCCTGATGTCGGGACAGCCCTGCTGTCAGATGAAAACA 331

Qy        491 GAGTGACCTGCACAAACTTGGAATTTTTTTTCCATACAACCCTGCCCCATCTACTGTA-- 548
              || ||||  | | || |  ||| |||||||| | |||||| | | | ||| | ||  |
Db        332 GAATGACACGTAAAATCCAGGATTTTTTTTTTC-TACAACTCCGACTCATTTGCTACATT 390

Qy        549 ---TTTTTTTTAATGAAATATGTGAATGATAAT 578
                  |||||||   |||||||||||||| |||||
Db        391 CCTTTTTTTTCTGTGAAATATGTGAATAATAAT 423


RESULT 11
US-09-918-995-32599
; Sequence 32599, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-756
;  CURRENT APPLICATION NUMBER: US/09/918,995
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US/09/235,076
;  PRIOR FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 38054
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32599
;   LENGTH: 457
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(457)
;   OTHER INFORMATION: n = A,T,C or G
US-09-918-995-32599

  Query Match             40.9%;  Score 238.2;  DB 9;  Length 457;
  Best Local Similarity   82.4%;  Pred. No. 5.9e-67;
  Matches  299;  Conservative    0;  Mismatches   58;  Indels    6;  Gaps    2;

Qy        221 CCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCC 280
              || |||| | || || ||| | | ||||||||| |  || ||| | || | |||||||||
Db         55 CCACTGCCTCTTCTCTTCTCGCTTGGGAACTCCAGTCTCACCTCGGCTTGCAATGGACCC 114

Qy        281 CAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGA 340
              ||||||||||||  | ||| ||||||||| ||||| |||||||||||||| |||||| |
Db        115 CAACTGCTCCTGCGAGGCTGGTGGCTCCTGCGCCTGCGCCGGCTCCTGCAAGTGCAAAAA 174

Qy        341 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAA 400
              |||||||||||||||||||||||||||||||||||| |||||  ||||||||||||||
Db        175 GTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGTTGCCCCCTGGGCTGTGCCAA 234

Qy        401 GTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATG 460
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        235 GTGTGCCCAGGGCTGCATCTGCAAAGGGGCGTCAGAGAAGTGCAGCTGCTGTGCCTGATG 294

Qy        461 TCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGAATTTTTTT 520
              || ||||||||||||   ||||  | | ||| ||||  | | || |  ||| |||||||
Db        295 TCGGGACAGCCCTGCTGTCAGATGAAAACAGAATGACACGTAAAATCCAGGATTTTTTTT 354

Qy        521 TCCATACAACCCTGCCCCATCTACTGTA-----TTTTTTTTAATGAAATATGTGAATGAT 575
              | | |||||| | | | ||| | ||  |     |||||||   |||||||||||||| ||
Db        355 TTC-TACAACTCCGACTCATTTGCTACATTCCTTTTTTTTTCTGTGAAATATGTGAATAAT 413

Qy        576 AAT 578
              |||
Db        414 AAT 416


RESULT 12
US-10-044-090-524
; Sequence 524, Application US/10044090
; Patent No. US20020137081A1
; GENERAL INFORMATION:
;  APPLICANT: Olga Bandman
;  TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
;  FILE REFERENCE: PA-0028 US
;  CURRENT APPLICATION NUMBER: US/10/044,090
;  CURRENT FILING DATE:  2002-01-09
;  NUMBER OF SEQ ID NOS: 850
;  SOFTWARE: PERL Program
; SEQ ID NO 524
;   LENGTH: 409
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20020137081A1 2717395CB1
US-10-044-090-524

  Query Match             40.3%;  Score 234.8;  DB 12;  Length 409;
  Best Local Similarity   80.3%;  Pred. No. 7.1e-66;
  Matches  302;  Conservative    0;  Mismatches   67;  Indels    7;  Gaps    2;

Qy        210 TCCAGCTCGAGCCGCTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTC 269
              ||||  |  | || |||| | ||  | ||| | | ||||||||  |  || ||| |  |
Db          3 TCCACGTGCACCCACTGCCTCTTCCCTTCTCGCTTGGGAACTCTAGTCTCGCCTCGGGTT 62

Qy        270 GAAATGGACCCCAACTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGC 329
              | |||||||||||||||||||  |  ||| || |||||  |||| ||| ||||||||
Db         63 GCAATGGACCCCAACTGCTCCTGTGCCGCTGGTGTCTCCTGCACCTGCGCCAGCTCCTGC 122

Qy        330 AAATGCAAAGAGTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTG 389
              || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 AAGTGCAAAGAGTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTG 182

Qy        390 GGCTGTGCCAAGTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGC 449
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        183 GGCTGTGCCAAGTGTGCCCAGGGCTGCATCTGCAAAGGGGCATCGGAGAAGTGCAGCTGC 242

Qy        450 TGTGCCTGATGTCCGGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTT 509
              || |||||||||| |||||||||||||| |  || | | ||||||||| | | || |
Db        243 TGCGCCTGATGTCGGGACAGCCCTGCTCCCAAGTACAAATAGAGTGACCCGTAAAATCCA 302

Qy        510 GGAATTT---TTTTTCCATACAACCCTGCCCCATCTACTGTA----TTTTTTTTAATGAA 562
              |||  |||   |||||   ||||| | || ||| | | ||  |    |||||||   ||||
Db        303 GGATTTTTTGTTTTTTGCTACAATCTTGACCCCTTTGCTACATTCCTTTTTTTTCTGTGAA 362

Qy        563 ATATGTGAATGATAAT 578
              |||||||||| |||||
Db        363 ATATGTGAATAATAAT 378


RESULT 13
US-10-044-090-835
; Sequence 835, Application US/10044090
; Patent No. US20020137081A1
; GENERAL INFORMATION:
;  APPLICANT: Olga Bandman
;  TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
;  FILE REFERENCE: PA-0028 US
;  CURRENT APPLICATION NUMBER: US/10/044,090
;  CURRENT FILING DATE:  2002-01-09
;  NUMBER OF SEQ ID NOS: 850
;  SOFTWARE: PERL Program
; SEQ ID NO 835
;   LENGTH: 523
;   TYPE: DNA
```

```
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20020137081A1 1568230CB1
US-10-044-090-835

  Query Match             40.2%;  Score 234.4;  DB 12;  Length 523;
  Best Local Similarity   82.1%;  Pred. No. 1.1e-65;
  Matches  294;  Conservative    0;  Mismatches   61;  Indels    3;  Gaps    2;

Qy        224 CTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCCCAA 283
              |||| ||||   |||      | || |||| || || ||||  ||||||||||||||||
Db        139 CTGCTTGTTCGTCTCACTGGTGTGAGCTCCAGCATCCCCTTTGCTCGAAATGGACCCCAA 198

Qy        284 CTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGAGTG 343
              |||||| ||| |  ||| |||||||||  | || ||||||||||||| ||||||||||
Db        199 CTGCTCTTGCGCCACTGGTGGCTCCTGCACGTGCGCCGGCTCCTGCAAGTGCAAAGAGTG 258

Qy        344 CAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAAGTG 403
              |||||||||||||||||||||||||||||| |||||||||| ||||||||||||||||
Db        259 CAAATGCACCTCCTGCAAGAAGAGCTGCTGTTCCTGCTGCCCCGTGGGCTGTGCCAAGTG 318

Qy        404 TGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATGTCC 463
              |||||||||||||| |||||||||||||||| ||||||||||||||||||||||||||||
Db        319 TGCCCAGGGCTGCGTCTGCAAAGGGGCATCGGAGAAGTGCAGCTGCTGTGCCTGATGTGG 378

Qy        464 GGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGAATTTTTTTTCC 523
              | ||||| || |||  |||| || | |||   |||||||| ||| |||| ||||||
Db        379 GAACAGCTCTTCTCCCAGATGTAAATAGAACAACCTGCAC-AACCTGGATTTTTTTAAAA 437

Qy        524 ATACAACCCTGCCCCATCTACTGTATTTTTTTTAAT--GAAATATGTGAATGATAATA 579
              ||||||| |||  |||| | ||| |||| |||| ||  ||||||||| ||| ||||
Db        438 ATACAACACTGAGCCATTTGCTGCATTTCTTTTTATACTAAATATGTGACTGACAATA 495


RESULT 14
US-09-925-300-821
; Sequence 821, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
;  APPLICANT: Craig Rosen,
;  APPLICANT:  Steve Ruben
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA101
;  CURRENT APPLICATION NUMBER: US/09/925,300
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05988
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1890
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 821
;   LENGTH: 453
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc feature
;   LOCATION: (29)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (392)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (409)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (430)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (433)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (434)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-925-300-821

  Query Match             39.9%;  Score 232.8;  DB 10;  Length 453;
  Best Local Similarity   81.8%;  Pred. No. 3.3e-65;
  Matches  292;  Conservative    0;  Mismatches   63;  Indels    2;  Gaps    2;

Qy        224 CTGCGTGTTTTCCTCTTGATCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCCCAA 283
              ||||  ||||   ||       | || |||| || || ||||  ||||||||||||||||
Db         15 CTGCTTGTTCGTCTNACTGGTGTGAGCTCCAGCATCCCCTTTGCTCGAAATGGACCCCAA 74

Qy        284 CTGCTCCTGCTCGCCTGTTGGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGAGTG 343
              |||||| ||| |  ||| |||||||||  | || ||||||||||||| ||||||||||
Db         75 CTGCTCTTGCGCCACTGGTGGCTCCTGCACGTGCGCCGGCTCCTGCAAGTGCAAAGAGTG 134

Qy        344 CAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAAGTG 403
              |||||||||||||||||||||||||||||| |||||||||| ||||||||||||||||
Db        135 CAAATGCACCTCCTGCAAGAAGAGCTGCTGTTCCTGCTGCCCCGTGGGCTGTGCCAAGTG 194

Qy        404 TGCCCAGGGCTGCATCTGCAAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATGTCC 463
              |||||||||||||| |||||||||||||||| ||||||||||||||||||||||||||||
Db        195 TGCCCAGGGCTGCGTCTGCAAAGGGGCATCGGAGAAGTGCAGCTGCTGTGCCTGATGTGG 254

Qy        464 GGACAGCCCTGCTCGAAGATATAGAAAGAGTGACCTGCACAAACTTGGAATTTTTTTTCC 523
              | ||||| || |||  | || || | |||   |||||||| ||| |||| ||||||
Db        255 GAACAGCTCTTCTCCCATATGTAAATAGAACAACCTGCAC-AACCTGGATTTTTTTAAAA 313

Qy        524 ATACAACCCTGCCCCATCTACTGTATTTTTTTTA-ATGAAATATGTGAATGATAATA 579
              ||||||| |||  |||| | ||| |||| ||||| |  |||||||||| ||| ||||
Db        314 ATACAACACTGAGCCATTTGCTGCATTTCTTTTATACTAAATATGTGACTGACAATA 370


RESULT 15
US-09-918-995-8793
; Sequence 8793, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-756
;  CURRENT APPLICATION NUMBER: US/09/918,995
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US/09/235,076
;  PRIOR FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 38054
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 8793
;   LENGTH: 409
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(409)
;   OTHER INFORMATION: n = A,T,C or G
US-09-918-995-8793

  Query Match             39.1%;  Score 228.2;  DB 9;  Length 409;
  Best Local Similarity   83.5%;  Pred. No. 9.8e-64;
  Matches  283;  Conservative    0;  Mismatches   53;  Indels    3;  Gaps    2;

Qy        243 TCGGGAACTCCTGCTTCTCCTTGCCTCGAAATGGACCCCAACTGCTCCTGCTCGCCTGTT 302
              | |||| |||| || || ||||  ||||||||||||||||||||||| ||| |  ||| |
Db          9 TGGGGAGCTCCAGCATCCCCTTTGCTCGAAATGGACCCCAACTGCTCTTGCGCCACTGGT 68

Qy        303 GGCTCCTGTGCCTGTGCCGGCTCCTGCAAATGCAAAGAGTGCAAATGCACCTCCTGCAAG 362
              ||||||||  | || ||||||| |||||| ||||||||||||||||||||||||||||
Db         69 GGCTCCTGCACGTGCGCCGGCTGCTGCAAGTGCAAAGAGTGCAAATGCACCTCCTGCAAG 128

Qy        363 AAGAGCTGCTGCTCCTGCTGCCCTGTGGGCTGTGCCAAGTGTGCCCAGGGCTGCATCTGC 422
              |||||||||||| |||||||||||| |||||||||||||||||||||||||||| |||||
Db        129 AAGAGCTGCTGTTCCTGCTGCCCCGTGGGCTGTGCCAAGTGTGCCCAGGGCTGCGTCTGC 188
```

```
Qy        423 AAAGGGGCATCAGAGAAGTGCAGCTGCTGTGCCTGATGTCCGGACAGCCCTGCTCGAAGA 482
              |||||||||||| |||||||||||||||||||||||||||||  | ||||| || |||  | |
Db        189 AAAGGGGCATCGGAGAAGTGCAGCTGCTGTGCCTGATGTGGGAACAGCTCTTCTCCCATA 248

Qy        483 TATAGAAAGAGTGACCTGCACAAACTTGGAATTTTTTTTCCATACAACCCTGCCCCATCT 542
              | || | |||   |||||||| ||| |||| ||||||    ||||||| |||  |||| |
Db        249 TGTAAATAGAACAACCTGCAC-AACCTGGATTTTTTTAAAAATACAACACTGAGCCATTT 307

Qy        543 ACTGTATTTTTTTTAAT--GAAATATGTGAATGATAATA 579
               ||| |||| |||| ||   |||||||||| ||| ||||
Db        308 GCTGCATTTCTTTTTATACTAAATATGTGACTGACAATA 346

Search completed: July 16, 2003, 18:04:19
Job time : 78.7139 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 16, 2003, 03:49:46 ; Search time 2975.96 Seconds
(without alignments)
15998.962 Million cell updates/sec

Title: US-09-981-353-165
Perfect score: 1636
Sequence: 1 gatcagtgtgtgagggaact.....ccaaattcaagaaagacctg 1636

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1636	100.0	1636	6	AX327326	AX327326 Sequence
2	1636	100.0	2759	6	AX395164	AX395164 Sequence
3	1636	100.0	2974	6	AX376454	AX376454 Sequence
4	1444.4	88.3	2786	9	AK025587	AK025587 Homo sapi
5	1020.4	62.4	2547	10	AF175221	AF175221 Cavia por
6	967.4	59.1	1650	6	AX040087	AX040087 Sequence
7	886	54.2	2103	10	BC025795	BC025795 Mus muscu
8	821.2	50.2	1766	9	HSA6054	AJ006054 Homo sapi
9	782	47.8	2090	6	AX410646	AX410646 Sequence
10	782	47.8	2090	9	HSU08854	U08854 Human UDP g
11	778.8	47.6	2114	9	AF180322	AF180322 Homo sapi
12	775.6	47.4	2107	6	AR168316	AR168316 Sequence
13	775.6	47.4	2107	9	HSU59209	U59209 Homo sapien
14	767.6	46.9	2108	9	AF112112	AF112112 Macaca fa
15	766.2	46.8	2123	6	AX410640	AX410640 Sequence
16	766.2	46.8	2123	9	HSU06641	U06641 Human UDP g
17	766	46.8	2079	9	HSAJ5162	AJ005162 Homo sapi
18	764.4	46.7	2129	9	BC026264	BC026264 Homo sapi
19	762.8	46.6	1584	10	AF184901	AF184901 Mus muscu
20	762.8	46.6	2097	9	AF064200	AF064200 Homo sapi
21	758	46.3	2091	9	AF081793	AF081793 Homo sapi
22	758	46.3	2093	6	AX411198	AX411198 Sequence
23	758	46.3	2093	9	HSUDPGT	Y00317 Human mRNA
24	757.6	46.3	2075	9	AF072223	AF072223 Macaca fa
25	754.8	46.1	1722	9	AF016492	AF016492 Homo sapi
26	754.8	46.1	1855	6	AX336329	AX336329 Sequence
27	754.8	46.1	1855	6	AX336696	AX336696 Sequence
28	754.8	46.1	1855	6	AX409473	AX409473 Sequence
29	754.8	46.1	1855	9	HUMUDPGTA	J05428 Human 3,4-c
30	751	45.9	1753	9	AF016310	AF016310 Macaca fa
31	750	45.8	1766	9	BC030974	BC030974 Homo sapi
32	748.4	45.7	2648	9	MFU91582	U91582 Macaca fasc
33	745.4	45.6	1915	9	AF401657	AF401657 Macaca fa
34	745.2	45.6	1662	9	AF294901	AF294901 Macaca mu
35	743.4	45.4	217991	9	AC021146	AC021146 Homo sapi
36	743	45.4	1831	10	RRUGTGENE	X57565 Rattus norv
37	738.8	45.2	1662	9	AF294902	AF294902 Macaca mu
38	736	45.0	1851	9	AF177272	AF177272 Homo sapi
39	734	44.9	1768	9	AF112113	AF112113 Macaca fa
40	732.6	44.8	2799	6	AX411112	AX411112 Sequence
41	732.6	44.8	2799	9	HSUGT2BIO	X63359 H.sapiens U
42	732.2	44.8	21000	6	AX395166	AX395166 Sequence
43	731.8	44.7	1671	4	RABUGT2BB	L01082 Oryctolagus
44	717.2	43.8	1825	10	AB034987	AB034987 Cavia por
45	701.4	42.9	1832	4	RABUGT2BA	L01081 Oryctolagus

ALIGNMENTS

RESULT 1
AX327326
LOCUS AX327326 1636 bp DNA linear PAT 07-JAN-2002
DEFINITION Sequence 11 from Patent WO0179468.
ACCESSION AX327326
VERSION AX327326.1 GI:18097872
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Policky,J.L., Hafalia,A., Burford,N., Ring,H.Z., Lal,P.,
Tribouley,C.M., Yao,M.G., Yue,H., Tang,Y.T., Patterson,C., Das,D.,
Sanjanwala,M.S., Gandhi,A.R., Reddy,R., Khan,F.A., Baughn,M.R.,

```
          Ramkumar,J., Griffin,J.A. and Au-Young,J.
  TITLE     Drug metabolizing enzymes
  JOURNAL   Patent: WO 0179468-A 11 25-OCT-2001;
            Incyte Genomics, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .1636
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /note="Incyte ID No: 2434655CB1"
BASE COUNT      483 a    315 c    369 g    469 t
ORIGIN
```

Query Match 100.0%; Score 1636; DB 6; Length 1636;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1636; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GATCAGTGTGTGAGGGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GATCAGTGTGTGAGGGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGC 60

Qy         61 TCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTG 120

Qy        121 ACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATG 180

Qy        181 AGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCAT 240

Qy        241 TGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTG 300

Qy        301 ACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATG 360

Qy        361 ATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATC 420

Qy        421 AGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGA 480

Qy        481 TTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAA 540

Qy        541 TTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATG 600

Qy        601 TACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATT 660

Qy        661 CAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAG 720

Qy        721 AGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTG 780

Qy        781 AGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACT 840

Qy        841 TTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAA 900

Qy        901 ATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTC 960

Qy        961 AAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 AAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGA 1020

Qy       1021 AGGTGTTATGGAGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGGTGTTATGGAGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGT 1080

Qy       1081 ATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTC 1140

Qy       1141 ATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTC 1200

Qy       1201 CCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAG 1260

Qy       1261 AAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTA 1320

Qy       1321 CCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTG 1380

Qy       1381 TAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCA 1440

Qy       1441 AGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGA 1500

Qy       1501 TTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTAT 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTAT 1560

Qy       1561 TTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1620

Qy       1621 ATTCAAGAAAGACCTG 1636
              ||||||||||||||||
Db       1621 ATTCAAGAAAGACCTG 1636
```

```
RESULT 2
AX395164
LOCUS       AX395164               2759 bp    DNA     linear   PAT 18-MAY-2002
DEFINITION  Sequence 1 from Patent WO0218554.
ACCESSION   AX395164
VERSION     AX395164.1  GI:21066217
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Wei,M.H., Ye,J., Ketchum,K.A., di Francesco,V. and Beasley,E.M.
  TITLE     Isolated human drug-metabolizing proteins, nucleic acid molecules
            encoding human drug-metabolizing proteins, and uses thereof
  JOURNAL   Patent: WO 0218554-A 1 07-MAR-2002;
            Applera Corporation (US)
```

```
FEATURES             Location/Qualifiers
     source          1. .2759
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      846 a    520 c    544 g    849 t
ORIGIN

  Query Match           100.0%;  Score 1636;  DB 6;  Length 2759;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1636;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GATCAGTGTGTGAGGGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         12 GATCAGTGTGTGAGGGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGC 71

Qy         61 TCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         72 TCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTG 131

Qy        121 ACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        132 ACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATG 191

Qy        181 AGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        192 AGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCAT 251

Qy        241 TGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        252 TGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTG 311

Qy        301 ACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        312 ACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATG 371

Qy        361 ATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        372 ATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATC 431

Qy        421 AGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        432 AGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGA 491

Qy        481 TTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        492 TTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAA 551

Qy        541 TTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        552 TTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATG 611

Qy        601 TACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        612 TACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATT 671

Qy        661 CAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        672 CAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAG 731

Qy        721 AGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        732 AGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTG 791

Qy        781 AGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        792 AGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACT 851

Qy        841 TTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        852 TTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAA 911

Qy        901 ATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        912 ATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTC 971

Qy        961 AAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        972 AAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGA 1031

Qy       1021 AGGTGTTATGGAGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1032 AGGTGTTATGGAGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGT 1091

Qy       1081 ATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1092 ATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTC 1151

Qy       1141 ATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1152 ATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTC 1211

Qy       1201 CCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1212 CCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAG 1271

Qy       1261 AAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1272 AAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTA 1331

Qy       1321 CCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1332 CCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTG 1391

Qy       1381 TAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1392 TAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCA 1451

Qy       1441 AGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1452 AGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGA 1511

Qy       1501 TTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTAT 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1512 TTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTAT 1571

Qy       1561 TTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1572 TTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1631

Qy       1621 ATTCAAGAAAGACCTG 1636
              ||||||||||||||||
Db       1632 ATTCAAGAAAGACCTG 1647


RESULT 3
AX376454
LOCUS       AX376454               2974 bp    DNA     linear   PAT 01-MAR-2002
DEFINITION  Sequence 521 from Patent WO0168848.
ACCESSION   AX376454
VERSION     AX376454.1  GI:19170617
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Baker,K.P., Chen,J., Desnoyers,L., Goddard,A., Godowski,P.J.,
            Gurney,A.L., Pan,J., Smith,V., Watanabe,C.K., Wood,W.I. and
            Zhang,Z.
  TITLE     Secreted and transmembrane polypeptides and nucleic acids encoding
            the same
  JOURNAL   Patent: WO 0168848-A 521 20-SEP-2001;
            Genentech, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .2974
                     /organism="Homo sapiens"
```

```
                     /db_xref="taxon:9606"
BASE COUNT      946 a    497 c    556 g    975 t
ORIGIN

  Query Match           100.0%;  Score 1636;  DB 6;  Length 2974;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1636;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GATCAGTGTGTGAGGGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGC  60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GATCAGTGTGTGAGGGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGC  60

Qy         61 TCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTG 120

Qy        121 ACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATG 180

Qy        181 AGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCAT 240

Qy        241 TGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTG 300

Qy        301 ACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATG 360

Qy        361 ATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATC 420

Qy        421 AGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGA 480

Qy        481 TTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAA 540

Qy        541 TTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATG 600

Qy        601 TACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATT 660

Qy        661 CAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAG 720

Qy        721 AGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTG 780

Qy        781 AGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACT 840

Qy        841 TTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAA 900

Qy        901 ATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTC 960

Qy        961 AAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 AAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGA 1020

Qy       1021 AGGTGTTATGGAGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGGTGTTATGGAGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGT 1080

Qy       1081 ATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTC 1140

Qy       1141 ATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTC 1200

Qy       1201 CCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAG 1260

Qy       1261 AAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTA 1320

Qy       1321 CCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTG 1380

Qy       1381 TAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCA 1440

Qy       1441 AGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGA 1500

Qy       1501 TTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTAT 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTAT 1560

Qy       1561 TTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1620

Qy       1621 ATTCAAGAAAGACCTG 1636
              ||||||||||||||||
Db       1621 ATTCAAGAAAGACCTG 1636


RESULT 4
AK025587
LOCUS       AK025587               2786 bp    mRNA    linear   PRI 29-SEP-2000
DEFINITION  Homo sapiens cDNA: FLJ21934 fis, clone HEP04364.
ACCESSION   AK025587
VERSION     AK025587.1  GI:10438147
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens hepatoma cell_line:HepG2 cDNA to mRNA, clone_lib:HEP
            clone:HEP04364.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
            Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
            Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2786)
  AUTHORS   Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
            Shibahara,T., Tanaka,T. and Nakamura,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,
            University of Tokyo, Laboratory of Genome Structure Analysis, Human
            Genome Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,
```

```
            Japan (E-mail:cdnal@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
            Fax:81-3-5449-5416)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing: Departent of
            Virology and Human Genome Center, Institute of Medical Science,
            University of Tokyo (partly supported by Science and Technology
            Agency).
FEATURES             Location/Qualifiers
     source          1. .2786
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HEP04364"
                     /cell_line="HepG2"
                     /cell_type="hepatoma"
                     /clone_lib="HEP"
                     /note="cloning vector pME18SFL3"
     CDS             71. .1420
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAB15179.1"
                     /db_xref="GI:10438148"
                     /translation="MPQDRTEENEIFVDLALNVLPGLSTWQSVIKLNDFFVEIRGTLK
                     MMCESFIYNQTLMKKLQETNYDVMLIDPVIPCGDLMAELLAVPFVLTLRISVGGNMER
                     SCGKLPAPLSYVPVPMTGLTDRMTFLERVKNSMLSVLFHFWIQDYDYHFWEEFYSKAL
                     GRPTTLCETVGKAEIWLIRTYWDFEFPQPYQPNFEFVGGLHCKPAKALPKEMENFVQS
                     SGEDGIVVFSLGSLFQNVTEEKANIIASALAQIPQKVLWRYKGKKPSTLGANTRLYDW
                     IPQNDLLGHPKTKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQLDNIAHMKAKGAAVE
                     INFKTMTSEDLLRALRTVITDSSYKENAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKG
                     AKHLRSAAHDLTWFQHYSIDVIGFLLACVATAIFLFTKCFLFSCQKFNKTRKIEKRE"
BASE COUNT      919 a    463 c    496 g    908 t
ORIGIN
```

Query Match 88.3%; Score 1444.4; DB 9; Length 2786;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1445; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      191 ATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCATTGAAATTTGA 250
Db        1 ATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCATTGAAATTTGA 60

Qy      251 GGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTGACCTAGCTCT 310
Db       61 GGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTGACCTAGCTCT 120

Qy      311 GAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATGATTTTTTTGT 370
Db      121 GAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATGATTTTTTTGT 180

Qy      371 TGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATCAGACGCTTAT 430
Db      181 TGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATCAGACGCTTAT 240

Qy      431 GAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGATTCCCTGTGG 490
Db      241 GAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGATTCCCTGTGG 300

Qy      491 AGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAATTTCTGTAGG 550
Db      301 AGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAATTTCTGTAGG 360

Qy      551 AGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATGTACCTGTGCC 610
Db      361 AGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATGTACCTGTGCC 420

Qy      611 TATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATTCAATGCTTTC 670
Db      421 TATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATTCAATGCTTTC 480

Qy      671 AGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAGAGTTTTATAG 730
Db      481 AGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAGAGTTTTATAG 540

Qy      731 TAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTGAGATATGGCT 790
Db      541 TAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTGAGATATGGCT 600

Qy      791 AATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACTTTGAGTTTGT 850
Db      601 AATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACTTTGAGTTTGT 660

Qy      851 TGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAAATTTTGTCCA 910
Db      661 TGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAAATTTTGTCCA 720

Qy      911 GAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTCAAAATGTTAC 970
Db      721 GAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTCAAAATGTTAC 780

Qy      971 AGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGAAGGTGTTATG 1030
Db      781 AGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGAAGGTGTTATG 840

Qy     1031 GAGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGTATGATTGGAT 1090
Db      841 GAGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGTATGATTGGAT 900

Qy     1091 ACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTCATGGTGGAAT 1150
Db      901 ACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTCATGGTGGAAT 960

Qy     1151 GAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTCCCATATTTGG 1210
Db      961 GAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTCCCATATTTGG 1020

Qy     1211 TGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAGAAATAAACTT 1270
Db     1021 TGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAGAAATAAACTT 1080

Qy     1271 CAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTACCGATTCCTC 1330
Db     1081 CAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTACCGATTCCTC 1140

Qy     1331 TTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTGTAAAGCCCCT 1390
Db     1141 TTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTGTAAAGCCCCT 1200

Qy     1391 AGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCAAGCACCTGCG 1450
Db     1201 AGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCAAGCACCTGCG 1260

Qy     1451 ATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGATTGGGTTCCT 1510
Db     1261 ATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGATTGGGTTCCT 1320

Qy     1511 GCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTATTTTCCTGTCA 1570
Db     1321 GCTGGCCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTATTTTCCTGTCA 1380

Qy     1571 AAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAAATTCAAGAAA 1630
Db     1381 AAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAAATTCAAGAAA 1440

Qy     1631 GACCTG 1636
Db     1441 GACCTG 1446
```

```
RESULT 5
AF175221
LOCUS       AF175221               2547 bp    mRNA    linear   ROD 18-MAY-2001
DEFINITION  Cavia porcellus UDP glucuronosyltransferase UGT2A3 mRNA, complete
            cds.
ACCESSION   AF175221
VERSION     AF175221.1  GI:5802603
```

```
KEYWORDS    .
SOURCE      Cavia porcellus.
  ORGANISM  Cavia porcellus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
REFERENCE   1  (bases 1 to 2547)
  AUTHORS   Smith,S.A., Nagalla,S.R., Andrews,D.P. and Olsen,G.D.
  TITLE     Morphine regulation of a novel uridine diphosphate
            glucuronosyl-transferase in guinea pig pups following in utero
            exposure
  JOURNAL   Mol. Genet. Metab. 68 (1), 68-77 (1999)
  MEDLINE   99410697
   PUBMED   10479484
REFERENCE   2  (bases 1 to 2547)
  AUTHORS   Smith,S.A., Nagalla,S.R., Andrews,D.P. and Olsen,G.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-1999) Pediatrics, Oregon Health Sciences
            University, 3181 SW Sam Jackson Park Road, Portland, OR 97201, USA
FEATURES             Location/Qualifiers
     source          1. .2547
                     /organism="Cavia porcellus"
                     /strain="Hartley"
                     /db_xref="taxon:10141"
                     /tissue_type="liver"
     CDS             20. .1612
                     /note="UGT"
                     /codon_start=1
                     /product="UDP glucuronosyltransferase UGT2A3"
                     /protein_id="AAD51732.1"
                     /db_xref="GI:5802604"
                     /translation="MAPGKLASAVLLLLLCCAGSGFCGKVLVWPCEMSHWLNLKTLLE
                     ELVKRGHEVTVLTLSNNLFIDYNRHPAFNFEVIPVPTDKNMSENILNEFIELAVNVMP
                     TMPLWQSGKLLQQFFVQITEDLGLNCRNTVYNQSLMKKLRDSKYDVLVTDPVIPCGEL
                     VAEMLGVPFVNMLKFSMGHTIEKYCGQLPAPPSYVPVPLGGLTTRMTFMERVKNMVFS
                     VLFDFWIQQYDYKFWDQFYSEALGRPTTLCEIMGKAEIWLIRTYWDFEFPRPYLPNFE
                     FVGGLHCKPAKPLPKEMEEFVQSSGEDGVVVFSLGSMVKNLTEEKANLIASALAQIPQ
                     KVLWRYKGKKPATLGPNTRLFDWIPQNDLLGHPKTKAFITHGGSNGIYEAIYHGVPMV
                     GMPIFSDQPDNLAGMKAKGAAVEVNMNTMTSADLLGALRTVINDPTYKENAMKLSRIH
                     HDQPVKPLDRAAFWVEFVMHHKGAKHLRVAAHDLSWFQYHSLDVIGFLLACVASAILL
                     VTKCCLFSFQNFIKIGKRIKKE"
BASE COUNT      765 a    469 c    523 g    790 t
ORIGIN

  Query Match             62.4%;  Score 1020.4;  DB 10;  Length 2547;
  Best Local Similarity   77.4%;  Pred. No. 2.7e-253;
  Matches 1253;  Conservative    0;  Mismatches  356;  Indels    9;  Gaps    1;

Qy         15 GGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGCTCCTGCAGCTCTTC 74
Db          8 GGAGAGGCCAGCATGGCTCCTGGGAAGTTGGCTTCTGCAGTCCTGCTGCTGCTGCTCTGC 67

Qy         75 TGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTGACATGAGCCATTGG 134
Db         68 TGTGCCGGCTCTGGATTCTGTGGGAAGGTGCTGGTGTGGCCGTGTGAAATGAGCCACTGG 127

Qy        135 CTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATGAGGTAACAGTATTG 194
Db        128 CTCAATTTAAAGACTCTTCTTGAGGAGCTTGTGAAAAGAGGACATGAGGTGACGGTTCTG 187

Qy        195 ACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCATTGAAATTTGAGGTG 254
Db        188 ACTCTTTCAAACAATTTGTTTATTGACTACAACAGACATCCTGCTTTTAATTTTGAAGTG 247

Qy        255 GTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATA---------TTTGTTGACCTA 305
Db        248 ATCCCTGTACCAACTGACAAAAACATGTCTGAGAATATACTAAATGAGTTTATAGAACTA 307

Qy        306 GCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATGATTTT 365
Db        308 GCTGTGAATGTCATGCCAACAATGCCTCTCTGGCAATCAGGAAAACTACTGCAACAATTC 367

Qy        366 TTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATCAGACG 425
Db        368 TTTGTTCAGATAACTGAAGACTTAGGACTTAACTGTAGAAATACAGTCTACAACCAGTCG 427

Qy        426 CTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGATTCCC 485
Db        428 CTCATGAAGAAGCTCCGGGACTCCAAGTATGATGTGTTGGTTACAGACCCTGTGATACCC 487

Qy        486 TGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAATTTCT 545
Db        488 TGTGGGGAGCTGGTGGCTGAGATGCTTGGGGTTCCTTTTGTGAACATGCTAAAGTTCTCC 547

Qy        546 GTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATGTACCT 605
Db        548 ATGGGTCATACCATAGAGAAATACTGTGGGCAGCTTCCAGCTCCACCTTCCTATGTGCCT 607

Qy        606 GTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATTCAATG 665
Db        608 GTTCCCCTGGGGGGACTCACAACCAGAATGACATTTATGGAAAGGGTGAAGAATATGGTG 667

Qy        666 CTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAGAGTTT 725
Db        668 TTTTCAGTTTTGTTTGACTTCTGGATCCAGCAGTATGACTATAAGTTTTGGGATCAGTTT 727

Qy        726 TATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTGAGATA 785
Db        728 TACAGTGAAGCTTTAGGAAGACCAACTACATTATGTGAGATTATGGGTAAAGCTGAAATT 787

Qy        786 TGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACTTTGAG 845
Db        788 TGGCTAATCCGGACATACTGGGATTTTGAATTTCCTCGTCCATATTTGCCTAATTTTGAA 847

Qy        846 TTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAAATTTT 905
Db        848 TTTGTTGGAGGACTGCACTGCAAACCTGCCAAACCTTTACCTAAGGAAATGGAAGAATTT 907

Qy        906 GTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTCAAAAT 965
Db        908 GTCCAGAGTTCAGGTGAAGATGGTGTGGTGGTGTTTTCTCTGGGGTCCATGGTGAAAAAT 967

Qy        966 GTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGAAGGTG 1025
Db        968 CTCACAGAGGAAAAAGCCAATCTCATTGCTTCAGCCCTTGCGCAGATTCCACAGAAGGTT 1027

Qy       1026 TTATGGAGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGTATGAT 1085
Db       1028 TTATGGAGATACAAGGGAAAGAAGCCAGCCACTTTAGGACCCAATACTCGGCTGTTTGAT 1087

Qy       1086 TGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTCATGGT 1145
Db       1088 TGGATTCCCCAGAATGATCTTCTTGGCCACCCCAAGACCAAAGCTTTCATCACTCATGGT 1147

Qy       1146 GGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTCCCATA 1205
Db       1148 GGATCTAATGGGATATATGAAGCTATTTATCATGGGGTCCCCATGGTGGGAATGCCCATA 1207

Qy       1206 TTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAGAAATA 1265
Db       1208 TTTTCTGATCAGCCTGACAACCTCGCTGGCATGAAGGCCAAAGGAGCAGCTGTGGAGGTC 1267

Qy       1266 AACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTACCGAT 1325
Db       1268 AACATGAACACAATGACAAGTGCAGACCTGCTGGGTGCCTTGAGGACAGTCATCAATGAC 1327

Qy       1326 TCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTGTAAAG 1385
Db       1328 CCTACTTATAAAGAGAATGCCATGAAGTTATCAAGAATTCACCATGACCAGCCAGTGAAG 1387

Qy       1386 CCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCAAGCAC 1445
Db       1388 CCCCTGGACCGAGCAGCTTTCTGGGTTGAGTTTGTCATGCACCATAAAGGGGCCAAGCAC 1447

Qy       1446 CTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGATTGGG 1505
Db       1448 CTGCGGGTTGCAGCCCATGACCTCTCCTGGTTCCAGTACCACTCTTTGGATGTGATTGGG 1507
```

```
Qy       1506 TTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTATTTTCC 1565
             ||||||||| ||||||| |||| ||||||||| || ||||||||||| ||||||||| 
Db       1508 TTCCTGCTGGCCTGTGTAGCAAGTGCTATATTGCTGGTCACAAAATGTTGCTTATTTTCC 1567

Qy       1566 TGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAAATT 1623
             | |||||| |||| ||| |   |||||| |  ||||| |||||||||| ||| |||||
Db       1568 TTTCAAAACTTTATTAAGATAGGAAAGAGAATAAAGAAGGAATAGATCATTCTAAATT 1625


RESULT 6
AX040087
LOCUS       AX040087               1650 bp    DNA     linear   PAT 18-NOV-2000
DEFINITION  Sequence 10 from Patent WO0063351.
ACCESSION   AX040087
VERSION     AX040087.1  GI:11230049
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1650)
  AUTHORS   Lal,P., Yue,H., Tang,Y.T., Hillman,J.L., Baughn,M.R. and Yang,J.
  TITLE     Carbohydrate-modifying enzymes
  JOURNAL   Patent: WO 0063351-A 10 26-OCT-2000;
            Incyte Genomics, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .1650
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /note="Incyte ID No: 2912330CB1"
BASE COUNT      489 a    330 c    354 g    477 t
ORIGIN

  Query Match           59.1%;  Score 967.4;  DB 6;  Length 1650;
  Best Local Similarity  75.7%;  Pred. No. 1.4e-239;
  Matches 1228;  Conservative    0;  Mismatches  386;  Indels    9;  Gaps    2;

Qy         23 CATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGCTCCTGCAGCTCTTCTGT---GT 79
              ||||| || ||||   || |   |||  |   |||| ||  | |||||   ||||    |
Db         28 CATCAGGATGTCTATGAAATGGACTTCAGCTCTTCTCCTGATACAGCTGAGCTGTTACTT 87

Qy         80 TGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTGACATGAGCCATTGGCTTAA 139
              | ||| |||    ||||| || || ||||||||||||   ||  | ||||| ||| | ||
Db         88 TAGCTCTGGGAGTTGTGGAAAGGTGCTGGTGTGGCCCACAGAATTCAGCCACTGGATGAA 147

Qy        140 TGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATGAGGTAACAGTATTGACTCA 199
              | | |||   || || || || ||  |   |||||| |||||||| || |||||| |
Db        148 TATAAAGACAATCCTGGATGAACTTGTCCAGAGAGGTCATGAGGTGACTGTATTGGCATC 207

Qy        200 CTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCATTGAAATTTGAGGTGGTCCA 259
              |||    |   |  | ||  ||  |  || ||| |  | |||||||| ||    |
Db        208 TTCAGCTTCCATTTCTTTCGATCCCAACAGCCCATCTACTCTTAAATTTGAAGTTTATCC 267

Qy        260 TATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTGACCTAGCTCTGA------A 313
              | |  |      | ||| ||     | || ||  |     | || | |  ||
Db        268 TGTATCTTTAACTAAAACTGAGTTTGAGGATATTATCAAGCAGCTGGTTAAGAGATGGGC 327

Qy        314 TGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATGATTTTTTTGTTGA 373
              |   | |||        ||   |||  ||   ||  | || || | ||  |  |
Db        328 AGAACTTCCAAAAGACACATTTTGGTCATATTTTTCACAAGTACAAGAAATCATGTGGAC 387

Qy        374 AATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATCAGACGCTTATGAA 433
              | | |  |   |  | | || | | ||| ||    |  | |  ||| |||  ||||||||
Db        388 ATTTAATGACATACTTAGAAAGTTCTGTAAGGATATAGTTTCAAATAAGAAACTTATGAA 447

Qy        434 GAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGATTCCCTGTGGAGA 493
              ||| ||||||||  | |  |  ||||| | |||  |||  |||| | |||| ||| ||
Db        448 GAAACTACAGGAGTCAAGATTTGATGTTGTTCTTGCAGATGCTGTTTTCCCCTTTGGTGA 507

Qy        494 CCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAATTTCTGTAGGAGG 553
              |||  |||| ||||| |||  | | || |||||   ||  ||  |  | |||   ||
Db        508 GCTGCTGGCCGAGTTACTTAAAATACCCTTTGTCTACAGCCTCCGCTTCTCTCCTGGCTA 567

Qy        554 CAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATGTACCTGTGCCTAT 613
              |   || ||        ||||  |||||     || | ||||||||| |||||  |||
Db        568 CGCAATTGAAAAGCATAGTGGAGGACTTCTGTTCCCTCCTTCCTATGTGCCTGTTGTTAT 627

Qy        614 GACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATTCAATGCTTTCAGT 673
              | ||| |||||  |||  |||||| ||  | || || |||||||||   ||  |   |
Db        628 GTCAGAACTAAGTGACCAAATGACTTTCATAGAGAGGGTAAAAAATATGATCTATGTGCT 687

Qy        674 TTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAGAGTTTTATAGTAA 733
              ||  ||  | || ||| | ||     |  |||     |    ||||| |||| || ||| |
Db        688 TTATTTTGAATTTTGGTTCCAAATATTTGACATGAAGAAGTGGGATCAGTTCTACAGTGA 747

Qy        734 GGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTGAGATATGGCTAAT 793
              |  |||||||  |||||||| |||| ||||||  ||| |||||||||  |||||||| ||
Db        748 AGTTCTAGGAAGACCCACTACGTTATCTGAGACAATGGCAAAAGCTGACATATGGCTTAT 807

Qy        794 ACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACTTTGAGTTTGTTGG 853
              ||||  || |||||||||  ||||||||| |||  |  ||| ||  |||||.|||||
Db        808 TCGAAACTACTGGGATTTTCAATTTCCTCACCCACTCTTACCAAATGTTGAGTTCGTTGG 867

Qy        854 AGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAAATTTTGTCCAGAG 913
              |||| | ||||| |||||||||||| |  | || ||||||||||| | ||||||||||
Db        868 AGGACTCCACTGCAAACCTGCCAAACCCCTACCGAAGGAAATGGAAGAGTTTGTCCAGAG 927

Qy        914 TTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTCAAAATGTTACAGA 973
              || || ||| ||||| ||||||||||||||||||||||  || |   ||     ||||
Db        928 CTCTGGAGAAAATGGTGTTGTGGTGTTTTCTCTGGGGTCGATGGTCAGTAACACGTCAGA 987

Qy        974 AGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGAAGGTGTTATGGAG 1033
              ||||| ||| ||| | ||||| |||||||||||| ||||||||| |||||  | |||||
Db        988 AGAAAGGGCCAATGTAATTGCATCAGCCCTTGCCAAGATCCCACAAAAGGTTCTGTGGAG 1047

Qy       1034 GTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGTATGATTGGATACC 1093
              |   | || || ||||||   || ||||||  |||||||||||||||  | ||||||||
Db       1048 ATTTGATGGGAATAAACCAGATACTTTAGGACTCAATACTCGGCTGTACAAGTGGATACC 1107

Qy       1094 CCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTCATGGTGGAATGAA 1153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1108 CCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTCATGGTGGAATGAA 1167

Qy       1154 TGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTCCCATATTTGGTGA 1213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1168 TGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTCCCATATTTGGTGA 1227

Qy       1214 TCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAGAAATAAACTTCAA 1273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1228 TCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAGAAATAAACTTCAA 1287

Qy       1274 AACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTACCGATTCCTCTTA 1333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1288 AACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTACCGATTCCTCTTA 1347

Qy       1334 TAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTGTAAAGCCCCTAGA 1393
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1348 TAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTGTAAAGCCCCTAGA 1407

Qy       1394 TCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCAAGCACCTGCGATC 1453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1408 TCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCAAGCACCTGCGATC 1467

Qy       1454 AGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGATTGGGTTCCTGCT 1513
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1468 AGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGATTGGGTTCCTGCT 1527

Qy       1514 GACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTATTTTCCTGTCAAAA 1573
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1528 GACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTATTTTCCTGTCAAAA 1587

Qy       1574 ATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAAATTCAAGAAAGAC 1633
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1588 ATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAAATTCAAGAAAGAC 1647
```

```
Qy       1634 CTG 1636
             |||
Db       1648 CTG 1650


RESULT 7
BC025795
LOCUS       BC025795               2103 bp    mRNA    linear   ROD 07-AUG-2002
DEFINITION  Mus musculus, RIKEN cDNA 2010321J07 gene, clone MGC:35622
            IMAGE:5133743, mRNA, complete cds.
ACCESSION   BC025795
VERSION     BC025795.1  GI:19387962
KEYWORDS    MGC.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2103)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAR-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
            Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
            Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
            George Yang, Scott Zuyderduyn, Marco Marra.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 57 Row: m Column: 15
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 13386283.
FEATURES             Location/Qualifiers
     source          1. .2103
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /map="FVB/N"
                     /clone="MGC:35622 IMAGE:5133743"
                     /tissue_type="Liver, normal. 5 month old male mouse."
                     /clone_lib="NCI_CGAP_Li9"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS             9. .1613
                     /codon_start=1
                     /product="RIKEN cDNA 2010321J07 gene"
                     /protein_id="AAH25795.1"
                     /db_xref="GI:19387963"
                     /db_xref="LocusID:72094"
                     /translation="MVSEKCVAAFFLLQLCWAGCGFCSKVLVWPCDMSHWLNLKTILE
                     ELGARGHEVTVLKYPSIIIDQSKRIPLHFENIPLLYEIETAENRLNEIANLAVNVIPN
                     LSLWEAAKTLQDFFLQVTGDFESICRSVLYNQKFMDKLRDAQYDVVVIDPVVPCGELV
                     AEVLQIPFVYTLRFSMGYYMEKHCGQLPIPLSYVPVVMSELTDNMTFTERVKNMMFSL
                     LFEYWLQQYDFAFWDQFYSETLGRPTTFCKTVGKADIWLIRTYWDVEFPRPYLPNFEF
                     VGGLHCKSAKPLPKEMEEFVQSSGEHGVVVFSLGSMVKNLTEEKANLIASVLAQIPQK
                     VLWRYSGKKPATLGSNTRLFNWIPQNDLLGHPKTKAFITHGGTNGIYEAIYHGVPMVG
                     VPMLGDQPHNIAHMEAKGAALKVSISTMTSTDLLSAVRAVINEPSYKENAMRLSRIHH
                     DQPVKPLDRAVFWIEFVMRHKGAKHLRVAAHDLSWFQYHSLDVIGFLLLCVVTLTFII
                     TKFCLFVCQKLYMKESKKMGNRKKKN"

BASE COUNT      626 a    408 c    461 g    608 t
ORIGIN

  Query Match           54.2%;  Score 886;  DB 10;  Length 2103;
  Best Local Similarity  73.5%;  Pred. No. 1.7e-218;
  Matches 1175;  Conservative    0;  Mismatches  410;  Indels   13;  Gaps    3;

Qy         23 CATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGCTCCTGCAGCTCTTCTGTGTTGG 82
              ||  |||   ||||| || |  | |  || |||| | || |||||||| | ||| |  ||
Db          5 CAATATGGTCTCTGAAAAATGTGTTGCGGCATTTTTTCTGCTGCAGCTTTGCTGGGCCGG 64

Qy         83 CTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTGACATGAGCCATTGGCTTAATGT 142
              ||||||||||||  | || ||||| ||||||||||||||| ||||||| ||||| ||| |
Db         65 CTGTGGATTCTGCAGCAAGGTCCTCGTGTGGCCCTGTGATATGAGCCACTGGCTGAATCT 124

Qy        143 CAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATGAGGTAACAGTATTGACTCACTC 202
              |||   ||||| || |||||   ||  ||||| || ||||||||||  |||   || |
Db        125 AAAGACTATTCTTGAGGAGCTTGGAGCAAGAGGGCACGAGGTAACAGTCCTGAAATACCC 184

Qy        203 AAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCATTGAAATTTGAGGTGGTCCATAT 262
              |    |    | || ||  | || || | |  | || || | ||||||    | | | |
Db        185 CA---GTATCATCATAGATCAGAGTAAACGTATTCCACTGCACTTTGAGAATATTCCTTT 241

Qy        263 GCCACAGGACAGAACAGAAGAAAATGAAATAT---------TTGTTGACCTAGCTCTGAA 313
              ||   | || |      ||   |  |      |||   | |||||| ||||
Db        242 GCTGTATGAAATCGAGACAGCTGAGAATCGTTTAAATGAGATTGCAAATCTAGCTGTGAA 301

Qy        314 TGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATGATTTTTTTGTTGA 373
              |||| | |||  | | |||  ||| || |||  | || |||| | || || ||| || |
Db        302 TGTCATTCCAAACCTGTCACTGTGGGAAGCAGCAAAAACATTACAAGACTTCTTTCTTCA 361

Qy        374 AATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATCAGACGCTTATGAA 433
              | |||  |||  |||  |||  || |||  |||  |  | ||||| ||||   | ||| |
Db        362 AGTAACTGGAGATTTTGAAAGTATTTGTAGGAGTGTATTGTACAACCAGAAATTCATGGA 421

Qy        434 GAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGATTCCCTGTGGAGA 493
              |||||||  |||  |  | || ||||||  || ||||||||||||  ||||||||||||
Db        422 CAAGCTACGGGATGCACAATATGATGTAGTGGTTATAGACCCTGTCGTTCCCTGTGGAGA 481

Qy        494 CCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAATTTCTGTAGGAGG 553
              || |||| ||  |||||     |||||| ||   ||||||  ||  |     | ||
Db        482 GTTGGTGGCAGAAGTGCTTCAGATCCCTTTCGTATACACACTGAGGTTCAGCATGGGCTA 541

Qy        554 CAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATGTACCTGTGCCTAT 613
              | | ||||||  |  ||||||  | ||||||  |||||| || |||||||| ||    ||
Db        542 CTACATGGAGAAACACTGTGGCCAGCTTCCAATTCCACTCTCGTATGTACCGGTTGTCAT 601

Qy        614 GACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATTCAATGCTTTCAGT 673
              ||  |  || |||||||  ||||||||    || || || ||||||   ||| ||||| |
Db        602 GAGTGAGCTGACAGACAATATGACCTTCACAGAGAGGGTGAAAAATATGATGTTTTCACT 661

Qy        674 TTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAGAGTTTTATAGTAA 733
              |||||  | | |||| | ||| | || |||| |   || |||||  |||||| ||| |
Db        662 GTTGTTTGAGTACTGGCTCCAGCAATATGACTTTGCATTCTGGGATCAGTTTTACAGTGA 721

Qy        734 GGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTGAGATATGGCTAAT 793
              |  |||||||||||||| || || ||| |||||||||| ||||||||  || ||||||||
Db        722 AACCCTAGGAAGGCCCACAACGTTCTGTAAGACTGTGGGGAAAGCTGACATTTGGCTAAT 781

Qy        794 ACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACTTTGAGTTTGTTGG 853
              ||||||||||||||| |||| |||||||  |||||   ||| || ||||||||||| ||
Db        782 CCGAACATATTGGGATGTTGAGTTTCCTCGTCCATATTTACCAAATTTTGAGTTTGTGGG 841

Qy        854 AGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAAATTTTGTCCAGAG 913
              |||| ||||||| ||| |||||||  |||| ||||||||||||||| | ||||| |||||
Db        842 AGGACTGCACTGCAAATCTGCCAAGCCTTTACCTAAGGAAATGGAAGAATTTGTTCAGAG 901

Qy        914 TTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTCAAAATGTTACAGA 973
              || || ||| ||||| | || || ||||| |||||||| || |  ||||  | |||||
Db        902 CTCTGGAGAACATGGTGTAGTAGTATTTTCACTGGGGTCAATGGTCAAAAACCTGACAGA 961

Qy        974 AGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGAAGGTGTTATGGAG 1033
              ||| || || ||  ||||||| || | |||||||||||| || |||||||| || |||||
```

```
Db        962 AGAGAAAGCCAACCTCATTGCCTCTGTCCTTGCCCAGATTCCCCAGAAGGTTTTGTGGAG 1021

Qy       1034 GTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGTATGATTGGATACC 1093

Db       1022 ATACTCAGGCAAGAAGCCAGCCACATTAGGATCCAATACTCGGCTTTTTAATTGGATTCC 1081

Qy       1094 CCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTCATGGTGGAATGAA 1153

Db       1082 CCAGAATGATCTTCTTGGACATCCTAAAACCAAAGCTTTCATCACACATGGTGGAACAAA 1141

Qy       1154 TGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTCCCATATTTGGTGA 1213

Db       1142 CGGGATTTATGAAGCCATTTACCATGGGGTCCCTATGGTGGGCGTTCCCATGTTAGGGGA 1201

Qy       1214 TCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAGAAATAAACTTCAA 1273

Db       1202 TCAGCCTCACAACATCGCTCACATGGAGGCCAAGGGAGCAGCCCTGAAAGTCAGCATCAG 1261

Qy       1274 AACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTACCGATTCCTCTTA 1333

Db       1262 TACAATGACGAGCACAGATTTACTCAGTGCTGTGAGGGCAGTCATTAATGAGCCTTCTTA 1321

Qy       1334 TAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTGTAAAGCCCCTAGA 1393

Db       1322 TAAAGAGAATGCCATGCGGTTATCAAGAATCCACCATGATCAGCCAGTGAAGCCCCTGGA 1381

Qy       1394 TCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCAAGCACCTGCGATC 1453

Db       1382 CCGAGCAGTCTTCTGGATTGAGTTTGTCATGCGTCACAAAGGAGCCAAGCATCTTCGTGT 1441

Qy       1454 AGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGATTGGGTTCCTGCT 1513

Db       1442 GGCAGCCCATGACCTCAGCTGGTTTCAGTACCACTCCCTAGATGTGATTGGGTTCCTATT 1501

Qy       1514 GACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTATTTTCCTGTCAAAA 1573

Db       1502 GTTGTGTGTCGTTACTCTGACATTCATCATCACTAAATTTTGTTTGTTTGTGTGTCAAAA 1561

Qy       1574 ATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGA 1611

Db       1562 ACTTTATATGA-AAGAAAGTAAGAAAATGGGGAACAGA 1598


RESULT 8
HSA6054
LOCUS       HSA6054                 1766 bp    mRNA    linear   PRI 16-JUL-1999
DEFINITION  Homo sapiens mRNA for UDP glucuronosyltransferase.
ACCESSION   AJ006054
VERSION     AJ006054.1  GI:4753765
KEYWORDS    UDP glucuronosyltransferase; ugt2A1 gene.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1766)
  AUTHORS   Jedlitschky,G., Cassidy,A.J., Sales,M., Pratt,N. and Burchell,B.
  TITLE     Cloning and characterization of a novel human olfactory
            UDP-glucuronosyltransferase
  JOURNAL   Biochem. J. 340 (Pt 3), 837-843 (1999)
  MEDLINE   99289328
   PUBMED   10359671
REFERENCE   2  (bases 1 to 1766)
  AUTHORS   Cassidy,A.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAY-1998) Cassidy A.J., Molecular and Cellular
            Pathology, University of Dundee, University Dept. of Molecular and
            Cellular pathology, Ninewells Hospital, Dundee, DD1 9SY, SCOTLAND
FEATURES             Location/Qualifiers
     source          1. .1766
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="olfactory"
     gene            64. .1647
                     /gene="ugt2A1"
     CDS             64. .1647
                     /gene="ugt2A1"
                     /EC_number="2.4.1.17"
                     /codon_start=1
                     /product="UDP glucuronosyltransferase"
                     /protein_id="CAB41974.1"
                     /db_xref="GI:4753766"
                     /translation="MLNNLLLFSLQISLIGTTLGGNVLIWPMEGSHWLNVKIIIDELI
                     KKEHNVTVLVASGALFITPTSNPSLTFEIYKVPFGKERIEGVIKDFVSTWLENRPSPS
                     TIWRFYQEMAKVIKDFHMVSQEICDGVLKNQQLMAKLKKSKFEVLVSDPVFPCGDIVA
                     LKLGIPFMYSLRFSPASTVEKHCGKVPYPPSYVPAVLSELTDQMSFTDRIRNFISYHL
                     QDYMFETLWKSWDSYYSKALGRPTTLCETMGKAEIWLIRTYWDFEFPRPYLPNFEFVG
                     GLHCKPAKPLPKEMEEFIQSSGKNGVVVFSLGSMVKNLTEEKANLIASALAQIPQKVL
                     WRYKGKKPATLGNNTQLFDWIPQNDLLGHPKTKAFITHGGTNGIYEAIYHGVPMVGVP
                     MFADQPDNIAHMKAKGAAVEVNLNTMTSVDLLSALRTVINEPSYKENAMRLSRIHHDQ
                     PVKPLDRAVFWIEFVMRHKGAKHLRVAAHDLTWFQYHSLDVIGFLLVCVTTAIFLVIQ
                     CCLFSCQKFGKIGKKKKRE"
BASE COUNT      542 a    359 c    363 g    502 t
ORIGIN

  Query Match           50.2%;  Score 821.2;  DB 9;  Length 1766;
  Best Local Similarity  71.9%;  Pred. No. 1e-201;
  Matches 1107;  Conservative    0;  Mismatches  418;  Indels   15;  Gaps    2;

Qy         94 GTGGGAAAGTCCTGGTGTGGCCCTGTGACATGAGCCATTGGCTTAATGTCAAGGTCATTC 153

Db        122 GTGGGAATGTTTTGATTTGGCCAATGGAAGGTAGTCATTGGCTAAATGTTAAGATAATTA 181

Qy        154 TAGAAGAGCTCATAGTGAGAGGCCATGAGGTAACAGTATTGACTCACTCAAAGCCTTCGT 213

Db        182 TAGATGAGCTCATTAAAAAGGAGCATAATGTGACTGTCCTAGTTGCCTCTGGTGCACTTT 241

Qy        214 TAATTGACTACAGGAAGCCTTCTGCATTGAAATTTGAGGTGGTCCATATGCCACAGGACA 273

Db        242 TCATCACACCAACCTCTAACCCATCTCTGACATTTGAAATATATAAGGTGCCCTTTGGCA 301

Qy        274 GAACAGAAGAAAATGAAATATTTGTTGACCTAGCTCTGAATGTCTTGCCAGGCTTATCAA 333

Db        302 AAGAAAGAATAGAAGGAGTAATTAAGGACTTCGTTTCGACATGGCTGGAAAATAGACCAT 361

Qy        334 CCTGGCAATCAGTTATAAAATTAAATGATTTTTTTGTTGAAATAAGAGGAACTTTAAAAA 393

Db        362 CTCCTTCAACCATTTGGAGATTCTATCAGGAGATGGCCAAAGTAATCAAGGACTTCCACA 421

Qy        394 TGATG------------TGTGAGAGCTTTATCTACAATCAGACGCTTATGAAGAAGCTAC 441

Db        422 TGGTGTCTCAGGAGATCTGTGATGGCGTTCTTAAAAACCAACAGCTGATGGCAAAGCTAA 481

Qy        442 AGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGATTCCCTGTGGAGACCTGATGG 501

Db        482 AGAAAAGCAAGTTTGAAGTCCTGGTGTCTGATCCAGTATTTCCTTGTGGCGATATAGTAG 541

Qy        502 CTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAATTTCTGTAGGAGGCAATATGG 561

Db        542 CTTTAAAACTTGGAATTCCATTTATGTACTCCTTGAGGTTTTCTCCAGCCTCAACAGTGG 601

Qy        562 AGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATGTACCTGTGCCTATGACAGGAC 621

Db        602 AAAAGCACTGTGGGAAGGTACCATACCCTCCTTCCTATGTTCCTGCTGTTTTATCAGAAC 661

Qy        622 TAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATTCAATGCTTTCAGTTTTGTTCC 681

Db        662 TCACCGACCAAATGTCTTTCACTGACAGAATAAGAAATT---TCATCTCCTACCACCTAC 718

Qy        682 ACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAGAGTTTTATAGTAAGGCATTAG 741

Db        719 AGGACTACATGTTTGAAACTCTTTGGAAATCATGGGATTCATACTATAGTAAAGCTTTAG 778

Qy        742 GAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTGAGATATGGCTAATACGAACAT 801

Db        779 GAAGACCCACTACGTTATGTGAGACTATGGGGAAAGCTGAAATTTGGTTAATCCGAACAT 838
```

```
Qy        802 ATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACTTTGAGTTTGTTGGAGGATTGC 861
Db        839 ATTGGGATTTTGAATTTCCTCGTCCATACTTACCTAATTTTGAGTTTGTTGGAGGATTGC 898

Qy        862 ACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAAATTTTGTCCAGAGTTCAGGGG 921
Db        899 ACTGCAAACCTGCCAAACCTTTACCTAAGGAAATGGAAGAATTTATCCAGAGCTCAGGTA 958

Qy        922 AAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTCAAAATGTTACAGAAGAAAAGG 981
Db        959 AAAATGGTGTTGTGGTGTTTTCTCTGGGATCAATGGTCAAAAACCTTACAGAAGAAAAGG 1018

Qy        982 CTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGAAGGTGTTATGGAGGTACAAAG 1041
Db       1019 CCAATCTTATTGCCTCAGCCCTTGCCCAGATTCCACAGAAGGTTTTATGGAGATACAAAG 1078

Qy       1042 GAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGTATGATTGGATACCCCAGAATG 1101
Db       1079 GAAAGAAACCAGCCACATTAGGAAACAATACTCAGCTCTTTGATTGGATACCCCAGAATG 1138

Qy       1102 ATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTCATGGTGGAATGAATGGGATCT 1161
Db       1139 ATCTTCTTGGACATCCCAAAACCAAAGCTTTTATCACTCATGGTGGAACTAATGGGATCT 1198

Qy       1162 ATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTCCCATATTTGGTGATCAGCTTG 1221
Db       1199 ACGAAGCTATTTACCACGGAGTCCCTATGGTGGGAGTTCCCATGTTTGCTGATCAGCCTG 1258

Qy       1222 ATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAGAAATAAACTTCAAAACTATGA 1281
Db       1259 ATAACATTGCTCACATGAAGGCCAAAGGAGCAGCTGTGGAAGTGAACCTAAACACAATGA 1318

Qy       1282 CAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTACCGATTCCTCTTATAAAGAGA 1341
Db       1319 CAAGTGTGGATTTGCTTAGCGCTTTGAGAACAGTCATTAATGAACCTTCTTATAAAGAGA 1378

Qy       1342 ATGCTATGAGATTATCAAGAATTCACCATGATCAACCTGTAAAGCCCCTAGATCGAGCAG 1401
Db       1379 ATGCTATGAGGTTATCAAGAATTCACCATGATCAACCTGTAAAGCCCCTGGATCGAGCAG 1438

Qy       1402 TCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCAAGCACCTGCGATCAGCTGCCC 1461
Db       1439 TCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCAAGCACCTTCGGGTTGCAGCCC 1498

Qy       1462 ATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGATTGGGTTCCTGCTGACCTGTG 1521
Db       1499 ATGACCTCACCTGGTTCCAGTACCACTCTTTGGATGTAATTGGGTTCTTGCTGGTCTGTG 1558

Qy       1522 TGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTATTTTCCTGTCAAAAATTTAATA 1581
Db       1559 TGACAACGGCTATATTTTTGGTCATACAATGTTGTTTGTTTTCCTGTCAAAAATTTGGTA 1618

Qy       1582 AAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAAA 1621
Db       1619 AGATAGGAAAGAAGAAAAAAAGAGAATAGGTCAAGAAAAA 1658
```

```
RESULT 9
AX410646
LOCUS       AX410646               2090 bp    DNA     linear   PAT 14-JUN-2002
DEFINITION  Sequence 3293 from Patent WO0229103.
ACCESSION   AX410646
VERSION     AX410646.1  GI:21443351
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Alvares,C., Horne,D., Peres-da-Silva,S. and Vockley,J.G.
  TITLE     Gene expression profiles in liver cancer
  JOURNAL   Patent: WO 0229103-A 3293 11-APR-2002;
            GENE LOGIC INC (US)
```

```
FEATURES             Location/Qualifiers
     source          1. .2090
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /note="EMBL/GenBank Accession No. U08854"
BASE COUNT      649 a    384 c    435 g    622 t
ORIGIN
```

Query Match 47.8%; Score 782; DB 6; Length 2090;
Best Local Similarity 68.8%; Pred. No. 1.5e-191;
Matches 1108; Conservative 0; Mismatches 490; Indels 12; Gaps 2;

```
Qy         24 ATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGCTCCTGCAGCTCTTCTGT---GTT 80
Db         16 ACCAGGATGTCTCTGAAATGGACGTCAGTCTTTCTGCTGATACAGCTCAGTTGTTACTTT 75

Qy         81 GGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTGACATGAGCCATTGGCTTAAT 140
Db         76 AGCTCTGGAAGCTGTGGAAAGGTGCTAGTGTGGCCCACAGAATACAGCCATTGGATAAAT 135

Qy        141 GTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATGAGGTAACAGTATTGACTCAC 200
Db        136 ATGAAGACAATCCTGGAAGAGCTTGTTCAGAGGGGTCATGAGGTGACTGTGTTGACATCT 195

Qy        201 TCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCATTGAAATTTGAGGTGGTCCAT 260
Db        196 TCGGCTTCTACTCTTGTCAATGCCAGTAAATCATCTGCTATTAAATTAGAAGTTTATCCT 255

Qy        261 ATGCC---------ACAGGACAGAACAGAAGAAAATGAAATATTTGTTGACCTAGCTCTG 311
Db        256 ACATCTTTAACTAAAAATGATTTGGAAGATTCTCTTCTGAAAATTCTCGATAGATGGATA 315

Qy        312 AATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATGATTTTTTTGTT 371
Db        316 TATGGTGTTTCAAAAAATACATTTTGGTCATATTTTTCACAATTACAAGAATTGTGTTGG 375

Qy        372 GAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATCAGACGCTTATG 431
Db        376 GAATATTATGACTACAGTAACAAGCTCTGTAAAGATGCAGTTTTGAATAAGAAACTTATG 435

Qy        432 AAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGATTCCCTGTGGA 491
Db        436 ATGAAACTACAAGAGTCAAAGTTTGATGTCATTCTGGCAGATGCCCTTAATCCCTGTGGT 495

Qy        492 GACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAATTTCTGTAGGA 551
Db        496 GAGCTACTGGCTGAACTATTTAACATACCCTTTCTGTACAGTCTTCGATTCTCTGTTGGC 555

Qy        552 GGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATGTACCTGTGCCT 611
Db        556 TACACATTTGAGAAGAATGGTGGAGGATTTCTGTTCCCTCCTTCCTATGTACCTGTTGTT 615

Qy        612 ATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATTCAATGCTTTCA 671
Db        616 ATGTCAGAATTAAGTGATCAAATGATTTTCATGGAGAGGATAAAAAATATGATACATATG 675

Qy        672 GTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAGAGTTTTATAGT 731
Db        676 CTTTATTTTGACTTTTGGTTTCAAATTTATGATCTGAAGAAGTGGGACCAGTTTTATAGT 735

Qy        732 AAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTGAGATATGGCTA 791
Db        736 GAAGTTCTAGGAAGACCCACTACATTATTTGAGACAATGGGGAAAGCTGAAATGTGGCTC 795

Qy        792 ATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACTTTGAGTTTGTT 851
Db        796 ATTCGAACCTATTGGGATTTTGAATTTCCTCGCCCATTCTTACCAAATGTTGATTTTGTT 855

Qy        852 GGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAAATTTTGTCCAG 911
Db        856 GGAGGACTTCACTGTAAACCAGCCAAACCCCTGCCTAAGGAAATGGAAGAGTTTGTGCAG 915

Qy        912 AGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTCAAAATGTTACA 971
```

```
Db        916 AGCTCTGGAGAAAATGGTATTGTGGTGTTTTCTCTGGGGTCGATGATCAGTAACATGTCA 975

Qy        972 GAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGAAGGTGTTATGG 1031

Db        976 GAAGAAAGTGCCAACATGATTGCATCAGCCCTTGCCCAGATCCCACAAAAGGTTCTATGG 1035

Qy       1032 AGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGTATGATTGGATA 1091

Db       1036 AGATTTGATGGCAAGAAGCCAAATACATTAGGTTCCAATACTCGACTGTACAAGTGGTTA 1095

Qy       1092 CCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTCATGGTGGAATG 1151

Db       1096 CCCCAGAATGACCTTCTTGGTCATCCCAAAACCAAAGCTTTTATAACTCATGGTGGAACC 1155

Qy       1152 AATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTCCCATATTTGGT 1211

Db       1156 AATGGCATCTATGAGGCGATCTACCATGGGATCCCTATGGTGGGCATTCCCTTGTTTGCG 1215

Qy       1212 GATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAGAAATAAACTTC 1271

Db       1216 GATCAACATGATAACATTGCTCACATGAAAGCCAAGGGAGCAGCCCTCAGTGTGGACATC 1275

Qy       1272 AAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTACCGATTCCTCT 1331

Db       1276 AGGACCATGTCAAGTAGAGATTTGCTCAATGCATTGAAGTCAGTCATTAATGACCCTGTC 1335

Qy       1332 TATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTGTAAAGCCCCTA 1391

Db       1336 TATAAAGAGAATGTCATGAAATTATCAAGAATTCATCATGACCAACCAATGAAGCCCCTG 1395

Qy       1392 GATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCAAGCACCTGCGA 1451

Db       1396 GATCGAGCAGTCTTCTGGATTGAGTTTGTCATGCGCCACAAAGGAGCCAAGCACCTTCGA 1455

Qy       1452 TCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGATTGGGTTCCTG 1511

Db       1456 GTCGCAGCTCACAACCTCACCTGGATCCAGTACCACTCTTTGGATGTGATAGCATTCCTG 1515

Qy       1512 CTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTATTTTCCTGTCAA 1571

Db       1516 CTGGCCTGCGTGGCAACTGTGATATTTATCATCACAAAATTTTGCCTGTTTTGTTTCCGA 1575

Qy       1572 AAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAAA 1621

Db       1576 AAGCTTGCCAAAACAGGAAAGAAGAAGAAAAGAGATTAGTTATATCAAAA 1625


RESULT 10
HSU08854
LOCUS       HSU08854                2090 bp    mRNA    linear   PRI 01-FEB-1995
DEFINITION  Human UDP glucuronosyltransferase precursor (UGT2B15) mRNA,
            complete cds.
ACCESSION   U08854
VERSION     U08854.1  GI:475758
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2090)
  AUTHORS   Green,M.D., Oturu,E.M. and Tephly,T.R.
  TITLE     Stable expression of a human liver UDP-glucuronosyltransferase
            (UGT2B15) with activity toward steroid and xenobiotic substrates
  JOURNAL   Drug Metab. Dispos. 22 (5), 799-805 (1994)
  MEDLINE   95136867
   PUBMED   7835232
REFERENCE   2  (bases 1 to 2090)
  AUTHORS   Green,M.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-APR-1994) Mitchell D. Green, Department of
            Pharmacology, The University of Iowa, Iowa City, IA 52242, USA
FEATURES             Location/Qualifiers
     source          1. .2090
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HE8a"
                     /tissue_type="liver"
                     /clone_lib="Human liver UniZap cDNA library from
                     Stratagene"
     gene            1. .2090
                     /gene="UGT2B15"
     CDS             22. .1614
                     /gene="UGT2B15"
                     /EC_number="2.4.1.17"
                     /codon_start=1
                     /product="UDP glucuronosyltransferase precursor"
                     /protein_id="AAC50077.1"
                     /db_xref="GI:475759"
                     /translation="MSLKWTSVFLLIQLSCYFSSGSCGKVLVWPTEYSHWINMKTILE
                     ELVQRGHEVTVLTSSASTLVNASKSSAIKLEVYPTSLTKNDLEDSLLKILDRWIYGVS
                     KNTFWSYFSQLQELCWEYYDYSNKLCKDAVLNKKLMMKLQESKFDVILADALNPCGEL
                     LAELFNIPFLYSLRFSVGYTFEKNGGGFLFPPSYVPVVMSELSDQMIFMERIKNMIHM
                     LYFDFWFQIYDLKKWDQFYSEVLGRPTTLFETMGKAEMWLIRTYWDFEFPRPFLPNVD
                     FVGGLHCKPAKPLPKEMEEFVQSSGENGIVVFSLGSMISNMSEESANMIASALAQIPQ
                     KVLWRFDGKKPNTLGSNTRLYKWLPQNDLLGHPKTKAFITHGGTNGIYEAIYHGIPMV
                     GIPLFADQHDNIAHMKAKGAALSVDIRTMSSRDLLNALKSVINDPVYKENVMKLSRIH
                     HDQPMKPLDRAVFWIEFVMRHKGAKHLRVAAHNLTWIQYHSLDVIAFLLACVATVIFI
                     ITKFCLFCFRKLAKTGKKKRD"
     sig_peptide     22. .90
                     /gene="UGT2B15"
     mat_peptide     91. .1611
                     /gene="UGT2B15"
                     /product="UDP glucuronosyltransferase"
                     /EC_number="2.4.1.17"
BASE COUNT      649 a    384 c    435 g    622 t
ORIGIN

  Query Match           47.8%;  Score 782;  DB 9;  Length 2090;
  Best Local Similarity  68.8%;  Pred. No. 1.5e-191;
  Matches 1108;  Conservative    0;  Mismatches  490;  Indels   12;  Gaps    2;

Qy         24 ATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGCTCCTGCAGCTCTTCTGT---GTT 80

Db         16 ACCAGGATGTCTCTGAAATGGACGTCAGTCTTTCTGCTGATACAGCTCAGTTGTTACTTT 75

Qy         81 GGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTGACATGAGCCATTGGCTTAAT 140

Db         76 AGCTCTGGAAGCTGTGGAAAGGTGCTAGTGTGGCCCACAGAATACAGCCATTGGATAAAT 135

Qy        141 GTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATGAGGTAACAGTATTGACTCAC 200

Db        136 ATGAAGACAATCCTGGAAGAGCTTGTTCAGAGGGGTCATGAGGTGACTGTGTTGACATCT 195

Qy        201 TCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCATTGAAATTTGAGGTGGTCCAT 260

Db        196 TCGGCTTCTACTCTTGTCAATGCCAGTAAATCATCTGCTATTAAATTAGAAGTTTATCCT 255

Qy        261 ATGCC---------ACAGGACAGAACAGAAGAAAATGAAATATTTGTTGACCTAGCTCTG 311

Db        256 ACATCTTTAACTAAAAATGATTTGGAAGATTCTCTTCTGAAAATTCTCGATAGATGGATA 315

Qy        312 AATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATGATTTTTTTGTT 371

Db        316 TATGGTGTTTCAAAAAATACATTTTGGTCATATTTTTCACAATTACAAGAATTGTGTTGG 375

Qy        372 GAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATCAGACGCTTATG 431

Db        376 GAATATTATGACTACAGTAACAAGCTCTGTAAAGATGCAGTTTTGAATAAGAAACTTATG 435

Qy        432 AAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGATTCCCTGTGGA 491

Db        436 ATGAAACTACAAGAGTCAAAGTTTGATGTCATTCTGGCAGATGCCCTTAATCCCTGTGGT 495

Qy        492 GACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAATTTCTGTAGGA 551
```

```
Db        496 GAGCTACTGGCTGAACTATTTAACATACCCTTTCTGTACAGTCTTCGATTCTCTGTTGGC 555

Qy        552 GGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATGTACCTGTGCCT 611
              ||   | |||   |   ||||   | |||     || | ||||||||||||||||   |
Db        556 TACACATTTGAGAAGAATGGTGGAGGATTTCTGTTCCCTCCTTCCTATGTACCTGTTGTT 615

Qy        612 ATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATTCAATGCTTTCA 671
              ||| ||| | |||  ||   |||||  ||  |||| ||  ||||||||   || | |
Db        616 ATGTCAGAATTAAGTGATCAAATGATTTTCATGGAGAGGATAAAAAATATGATACATATG 675

Qy        672 GTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAGAGTTTTATAGT 731
              |||  ||  |||| ||| ||||   ||| ||     |    |||||  ||||||||||
Db        676 CTTTATTTTGACTTTTGGTTTCAAATTTATGATCTGAAGAAGTGGGACCAGTTTTATAGT 735

Qy        732 AAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTGAGATATGGCTA 791
              | |  ||||||| |||||||||||||| ||||||  |||| ||||||| || ||||||
Db        736 GAAGTTCTAGGAAGACCCACTACATTATTTGAGACAATGGGGAAAGCTGAAATGTGGCTC 795

Qy        792 ATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACTTTGAGTTTGTT 851
              || ||||| |||||||||||||||||||||  |||| |  ||| ||  |||| ||||||
Db        796 ATTCGAACCTATTGGGATTTTGAATTTCCTCGCCCATTCTTACCAAATGTTGATTTTGTT 855

Qy        852 GGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAAATTTTGTCCAG 911
              |||||| | |||||||||| |||||| |  |||||||||||||||| | ||||| |||
Db        856 GGAGGACTTCACTGTAAACCAGCCAAACCCCTGCCTAAGGAAATGGAAGAGTTTGTGCAG 915

Qy        912 AGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTCAAAATGTTACA 971
              || || || ||| ||||||||||||||||||||||||||  || |   ||  |  ||
Db        916 AGCTCTGGAGAAAATGGTATTGTGGTGTTTTCTCTGGGGTCGATGATCAGTAACATGTCA 975

Qy        972 GAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGAAGGTGTTATGG 1031
              |||||||  || || || ||||| ||||||||||||||||||||| |||||  |||||
Db        976 GAAGAAAGTGCCAACATGATTGCATCAGCCCTTGCCCAGATCCCACAAAAGGTTCTATGG 1035

Qy       1032 AGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGTATGATTGGATA 1091
              || |   | || || || |||   ||||||||  |||||||||| |||||  | ||| ||
Db       1036 AGATTTGATGGCAAGAAGCCAAATACATTAGGTTCCAATACTCGACTGTACAAGTGGTTA 1095

Qy       1092 CCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTCATGGTGGAATG 1151
              |||||||||| |||||||||||||||||||||||||||||||||| ||||||||||||
Db       1096 CCCCAGAATGACCTTCTTGGTCATCCCAAAACCAAAGCTTTTATAACTCATGGTGGAACC 1155

Qy       1152 AATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTCCCATATTTGGT 1211
              ||||| |||||||| || || |||||||| |||||||||||  ||||| | ||||
Db       1156 AATGGCATCTATGAGGCGATCTACCATGGGATCCCTATGGTGGGCATTCCCTTGTTTGCG 1215

Qy       1212 GATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAGAAATAAACTTC 1271
              ||||| | |||||||| |||||||||| ||||| ||||||||  |     |  || ||
Db       1216 GATCAACATGATAACATTGCTCACATGAAAGCCAAGGGAGCAGCCCTCAGTGTGGACATC 1275

Qy       1272 AAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTACCGATTCCTCT 1331
              |  || ||| ||||   ||||| || |  || ||||   |||||||||  ||  | .
Db       1276 AGGACCATGTCAAGTAGAGATTTGCTCAATGCATTGAAGTCAGTCATTAATGACCCTGTC 1335

Qy       1332 TATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTGTAAAGCCCCTA 1391
              ||||||||||||  |||| ||||||||||||||| ||||| ||||| | ||||||||
Db       1336 TATAAAGAGAATGTCATGAAATTATCAAGAATTCATCATGACCAACCAATGAAGCCCCTG 1395

Qy       1392 GATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCAAGCACCTGCGA 1451
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||| |||
Db       1396 GATCGAGCAGTCTTCTGGATTGAGTTTGTCATGCGCCACAAAGGAGCCAAGCACCTTCGA 1455

Qy       1452 TCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGATTGGGTTCCTG 1511
                 || ||  |||||||||| |||| || ||||| | |||||||  |  ||||||
Db       1456 GTCGCAGCTCACAACCTCACCTGGATCCAGTACCACTCTTTGGATGTGATAGCATTCCTG 1515

Qy       1512 CTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTATTTTCCTGTCAA 1571
              ||| |||| ||||||||||  |||||  |  |||||||| ||    | ||||  |  | |
Db       1516 CTGGCCTGCGTGGCAACTGTGATATTTATCATCACAAAATTTTGCCTGTTTTGTTTCCGA 1575

Qy       1572 AAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAAA 1621
              ||  ||   |||||  ||||||   | || || || ||| | | || |||
Db       1576 AAGCTTGCCAAAACAGGAAAGAAGAAGAAAAGAGATTAGTTATATCAAAA 1625
```

```
RESULT 11
AF180322
LOCUS       AF180322               2114 bp    mRNA    linear   PRI 14-SEP-1999
DEFINITION  Homo sapiens UDP-glucuronosyltransferase 2B15 (UGT2B15) mRNA,
            UGT2B15-Y85 allele, complete cds.
ACCESSION   AF180322
VERSION     AF180322.1  GI:5881245
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2114)
  AUTHORS   Levesque,E., Beaulieu,M., Green,M.D., Tephly,T.R., Belanger,A. and
            Hum,D.W.
  TITLE     Isolation and characterization of UGT2B15(Y85): a
            UDP-glucuronosyltransferase encoded by a polymorphic gene
  JOURNAL   Pharmacogenetics 7 (4), 317-325 (1997)
  MEDLINE   97439504
   PUBMED   9295060
REFERENCE   2  (bases 1 to 2114)
  AUTHORS   Levesque,E., Beaulieu,M., Green,M.D., Tephly,T.R., Belanger,A. and
            Hum,D.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-AUG-1999) Molecular Endocrinology, CHUL Research
            Center, 2705 Laurier Blvd, Ste-Foy, Quebec G1V 4G2, Canada
FEATURES             Location/Qualifiers
     source          1. .2114
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4q13"
     gene            1. .2114
                     /gene="UGT2B15"
                     /allele="Y85"
     CDS             39. .1631
                     /gene="UGT2B15"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase 2B15"
                     /protein_id="AAD55093.1"
                     /db_xref="GI:5881246"
                     /translation="MSLKWTSVFLLIQLSCYFSSGSCGKVLVWPTEYSHWINMKTILE
                     ELVQRGHEVTVLTSSASTLVNASKSSAIKLEVYPTSLTKNYLEDSLLKILDRWIYGVS
                     KNTFWSYFSQLQELCWEYYDYSNKLCKDAVLNKKLMMKLQESKFDVILADALNPCGEL
                     LAELFNIPFLYSLRFSVGYTFEKNGGGFLFPPSYVPVVMSELSDQMIFMERIKNMIHM
                     LYFDFWFQIYDLKKWDQFYSEVLGRPTTLFETMGKAEMWLIRTYWDFEFPRPFLPNVD
                     FVGGLHCKPAKPLPKEMEEFVQSSGENGIVVFSLGSMISNMSEESANMIASALAQIPQ
                     KVLWRFDGKKPNTLGSNTRLYKWLPQNDLLGHPKTKAFITHGGTNGIYEAIYHGIPMV
                     GIPLFADQHDNIAHMKAKGAALSVDIRTMSSRDLLNALKSVINDPVYKENVMKLSRIH
                     HDQPMKPLDRAVFWIEFVMRHKGAKHLRVAAHNLTWIQYHSLDVIAFLLACVATVIFI
                     ITKFCLFCFRKLAKTGKKKRD"
BASE COUNT      670 a    386 c    435 g    623 t
ORIGIN

  Query Match           47.6%;  Score 778.8;  DB 9;  Length 2114;
  Best Local Similarity   68.7%;  Pred. No. 1e-190;
  Matches 1106;  Conservative    0;  Mismatches  492;  Indels   12;  Gaps    2;

Qy         24 ATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGCTCCTGCAGCTCTTCTGT---GTT 80
              | || || ||||   || |   | |  || ||||||||  | ||||||   |||    ||
Db         33 ACCAGGATGTCTCTGAAATGGACGTCAGTCTTTCTGCTGATACAGCTCAGTTGTTACTTT 92

Qy         81 GGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTGACATGAGCCATTGGCTTAAT 140
              |||  ||||  |||||| || || || |||||||||   ||    |||||||| | |||
Db         93 AGCTCTGGAAGCTGTGGAAAGGTGCTAGTGTGGCCCACAGAATACAGCCATTGGATAAAT 152

Qy        141 GTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATGAGGTAACAGTATTGACTCAC 200
              | |||   || || ||||||||  |  ,||| || |||||||| || || |||||
Db        153 ATGAAGACAATCCTGGAAGAGCTTGTTCAGAGGGGTCATGAGGTGACTGTGTTGACATCT 212

Qy        201 TCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCATTGAAATTTGAGGTGGTCCAT 260
```

```
Db        213 TCGGCTTCTACTCTTGTCAATGCCAGTAAATCATCTGCTATTAAATTAGAAGTTTATCCT 272
Qy        261 ATGCC---------ACAGGACAGAACAGAAGAAAATGAAATATTTGTTGACCTAGCTCTG 311
Db        273 ACATCTTTAACTAAAAATTATTTGGAAGATTCTCTTCTGAAAATTCTCGATAGATGGATA 332
Qy        312 AATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATGATTTTTTTGTT 371
Db        333 TATGGTGTTTCAAAAAATACATTTTGGTCATATTTTTCACAATTACAAGAATTGTGTTGG 392
Qy        372 GAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATCAGACGCTTATG 431
Db        393 GAATATTATGACTACAGTAACAAGCTCTGTAAAGATGCAGTTTTGAATAAGAAACTTATG 452
Qy        432 AAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGATTCCCTGTGGA 491
Db        453 ATGAAACTACAAGAGTCAAAGTTTGATGTCATTCTGGCAGATGCCCTTAATCCCTGTGGT 512
Qy        492 GACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAATTTCTGTAGGA 551
Db        513 GAGCTACTGGCTGAACTATTTAACATACCCTTTCTGTACAGTCTTCGATTCTCTGTTGGC 572
Qy        552 GGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATGTACCTGTGCCT 611
Db        573 TACACATTTGAGAAGAATGGTGGAGGATTTCTGTTCCCTCCTTCCTATGTACCTGTTGTT 632
Qy        612 ATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATTCAATGCTTTCA 671
Db        633 ATGTCAGAATTAAGTGATCAAATGATTTTCATGGAGAGGATAAAAAATATGATACATATG 692
Qy        672 GTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAGAGTTTTATAGT 731
Db        693 CTTTATTTTGACTTTTGGTTTCAAATTTATGATCTGAAGAAGTGGGACCAGTTTTATAGT 752
Qy        732 AAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTGAGATATGGCTA 791
Db        753 GAAGTTCTAGGAAGACCCACTACATTATTTGAGACAATGGGGAAAGCTGAAATGTGGCTC 812
Qy        792 ATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACTTTGAGTTTGTT 851
Db        813 ATTCGAACCTATTGGGATTTTGAATTTCCTCGCCCATTCTTACCAAATGTTGATTTTGTT 872
Qy        852 GGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAAATTTTGTCCAG 911
Db        873 GGAGGACTTCACTGTAAACCAGCCAAACCCCTGCCTAAGGAAATGGAAGAGTTTGTGCAG 932
Qy        912 AGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTCAAAATGTTACA 971
Db        933 AGCTCTGGAGAAAATGGTATTGTGGTGTTTTCTCTGGGGTCGATGATCAGTAACATGTCA 992
Qy        972 GAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGAAGGTGTTATGG 1031
Db        993 GAAGAAAGTGCCAACATGATTGCATCAGCCCTTGCCCAGATCCCACAAAAGGTTCTATGG 1052
Qy       1032 AGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGTATGATTGGATA 1091
Db       1053 AGATTTGATGGCAAGAAGCCAAATACTTTAGGTTCCAATACTCGACTGTACAAGTGGTTA 1112
Qy       1092 CCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTCATGGTGGAATG 1151
Db       1113 CCCCAGAATGACCTTCTTGGTCATCCCAAAACCAAAGCTTTTATAACTCATGGTGGAACC 1172
Qy       1152 AATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTCCCATATTTGGT 1211
Db       1173 AATGGCATCTATGAGGCGATCTACCATGGGATCCCTATGGTGGGCATTCCCTTGTTTGCG 1232
Qy       1212 GATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAGAAATAAACTTC 1271
Db       1233 GATCAACATGATAACATTGCTCACATGAAAGCCAAGGGAGCAGCCCTCAGTGTGGACATC 1292
Qy       1272 AAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTACCGATTCCTCT 1331
```

```
Db        1293 AGGACCATGTCAAGTAGAGATTTGCTCAATGCATTGAAGTCAGTCATTAATGACCCTGTC 1352
Qy        1332 TATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTGTAAAGCCCCTA 1391
Db        1353 TATAAAGAGAATGTCATGAAATTATCAAGAATTCATCATGACCAACCAATGAAGCCCCTG 1412
Qy        1392 GATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCAAGCACCTGCGA 1451
Db        1413 GATCGAGCAGTCTTCTGGATTGAGTTTGTCATGCGCCACAAAGGAGCCAAGCACCTTCGA 1472
Qy        1452 TCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGATTGGGTTCCTG 1511
Db        1473 GTCGCAGCTCACAACCTCACCTGGATCCAGTACCACTCTTTGGATGTGATAGCATTCCTG 1532
Qy        1512 CTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTATTTTCCTGTCAA 1571
Db        1533 CTGGCCTGCGTGGCAACTGTGATATTTATCATCACAAAATTTTGCCTGTTTTGTTTCCGA 1592
Qy        1572 AAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAAA 1621
Db        1593 AAGCTTGCCAAAACAGGAAAGAAGAAGAAAAGAGATTAGTTATATCAAAA 1642
```

RESULT 12
AR168316
LOCUS AR168316 2107 bp DNA linear PAT 17-DEC-2001
DEFINITION Sequence 1 from patent US 6287834.
ACCESSION AR168316
VERSION AR168316.1 GI:17904166
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 2107)
AUTHORS Belanger,A., Hum,D.W., Beaulieu,M. and Levesque,E.
TITLE Characterization and use of an isolated uridine diphospho-glucuronosyltransferase
JOURNAL Patent: US 6287834-A 1 11-SEP-2001;
FEATURES Location/Qualifiers
source 1. .2107
/organism="unknown"
BASE COUNT 657 a 382 c 424 g 644 t
ORIGIN

Query Match 47.4%; Score 775.6; DB 6; Length 2107;
Best Local Similarity 68.6%; Pred. No. 6.8e-190;
Matches 1104; Conservative 0; Mismatches 494; Indels 12; Gaps 2;

```
Qy          24 ATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGCTCCTGCAGCTCTTCTGT---GTT 80
Db          46 ACCAGGATGTCTCTGAAATGGATGTCAGTCTTTCTGCTGATGCAGCTCAGTTGTTACTTT 105
Qy          81 GGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTGACATGAGCCATTGGCTTAAT 140
Db         106 AGCTCTGGGAGTTGTGGAAAGGTGCTGGTGTGGCCCACAGAATACAGCCATTGGATAAAT 165
Qy         141 GTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATGAGGTAACAGTATTGACTCAC 200
Db         166 ATGAAGACAATCCTGGAAGAGCTTGTTCAGAGGGGTCATGAGGTGATTGTGTTGACATCT 225
Qy         201 TCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCATTGAAATTTGAGGTGGTCCAT 260
Db         226 TCGGCTTCTATTCTTGTCAATGCCAGTAAATCATCTGCTATTAAATTAGAAGTTTATCCT 285
Qy         261 ATGCCACAGGACAGAACAGAAGAAAATGAAATATTT---------GTTGACCTAGCTCTG 311
Db         286 ACATCTTTAACTAAAAATGATTTGGAAGATTTTTTTATGAAAATGTTCGATAGATGGACA 345
Qy         312 AATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATGATTTTTTTGTT 371
Db         346 TATAGTATTTCAAAAAATACATTTTGGTCATATTTTTCACAACTACAAGAATTGTGTTGG 405
Qy         372 GAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATCAGACGCTTATG 431
```

```
Db        406 GAATATTCTGACTATAATATAAAGCTCTGTGAAGATGCAGTTTTGAACAAGAAACTTATG 465

Qy        432 AAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGATTCCCTGTGGA 491

Db        466 AGAAAACTACAAGAGTCAAAATTTGATGTCCTTCTGGCAGATGCCGTTAATCCCTGTGGT 525

Qy        492 GACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAATTTCTGTAGGA 551

Db        526 GAGCTGCTGGCTGAACTACTTAACATACCCTTTCTGTACAGTCTCCGCTTCTCTGTTGGC 585

Qy        552 GGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATGTACCTGTGCCT 611

Db        586 TACACAGTTGAGAAGAATGGTGGAGGATTTCTGTTCCCTCCTTCCTATGTACCTGTTGTT 645

Qy        612 ATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATTCAATGCTTTCA 671

Db        646 ATGTCAGAATTAAGTGATCAAATGATTTTCATGGAGAGGATAAAAAATATGATATATATG 705

Qy        672 GTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAGAGTTTTATAGT 731

Db        706 CTTTATTTTGACTTTTGGTTTCAAGCATATGATCTGAAGAAGTGGGACCAGTTTTATAGT 765

Qy        732 AAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTGAGATATGGCTA 791

Db        766 GAAGTTCTAGGAAGACCCACTACATTATTTGAGACAATGGGGAAAGCTGAAATGTGGCTC 825

Qy        792 ATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACTTTGAGTTTGTT 851

Db        826 ATTCGAACCTATTGGGATTTTGAATTTCCTCGCCCATTCTTACCAAATGTTGATTTTGTT 885

Qy        852 GGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAAATTTTGTCCAG 911

Db        886 GGAGGACTTCACTGTAAACCAGCCAAACCCTTGCCTAAGGAAATGGAAGAGTTTGTGCAG 945

Qy        912 AGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTCAAAATGTTACA 971

Db        946 AGCTCTGGAGAAAATGGTATTGTGGTGTTTTCTCTGGGGTCGATGATCAGTAACATGTCA 1005

Qy        972 GAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGAAGGTGTTATGG 1031

Db       1006 GAAGAAAGTGCCAACATGATTGCATCAGCCCTTGCCCAGATCCCACAAAAGGTTCTATGG 1065

Qy       1032 AGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGTATGATTGGATA 1091

Db       1066 AGATTTGATGGCAAGAAGCCAAATACTTTAGGTTCCAATACTCGACTGTATAAGTGGTTA 1125

Qy       1092 CCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTCATGGTGGAATG 1151

Db       1126 CCCCAGAATGACCTTCTTGGTCATCCCAAAACCAAAGCTTTTATAACTCATGGTGGAACC 1185

Qy       1152 AATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTCCCATATTTGGT 1211

Db       1186 AATGGCATCTATGAGGCGATCTACCATGGGATCCCTATGGTGGGCATTCCCTTGTTTGCG 1245

Qy       1212 GATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAGAAATAAACTTC 1271

Db       1246 GATCAACATGATAACATTGCTCACATGAAAGCCAAGGGAGCAGCCCTCAGTGTGGACATC 1305

Qy       1272 AAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTACCGATTCCTCT 1331

Db       1306 AGGACCATGTCAAGTAGAGATTTGCTCAATGCATTGAAGTCAGTCATTAATGACCCTATC 1365

Qy       1332 TATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTGTAAAGCCCCTA 1391

Db       1366 TATAAAGAGAATATCATGAAATTATCAAGAATTCATCATGATCAACCGGTGAAGCCCCTG 1425

Qy       1392 GATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCAAGCACCTGCGA 1451

Db       1426 GATCGAGCAGTCTTCTGGATTGAGTTTGTCATGCGCCATAAAGGAGCCAAGCACCTTCGG 1485

Qy       1452 TCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGATTGGGTTCCTG 1511

Db       1486 GTCGCAGCCCACAACCTCACCTGGATCCAGTACCACTCTTTGGATGTGATAGCATTCCTG 1545

Qy       1512 CTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTATTTTCCTGTCAA 1571

Db       1546 CTGGCCTGCGTGGCAACTATGATATTTATGATCACAAAATGTTGCCTGTTTTGTTTCCGA 1605

Qy       1572 AAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAAA 1621

Db       1606 AAGCTTGCCAAAACAGGAAAGAAGAAGAAAAGGGATTAGTTATATCAAAA 1655
```

RESULT 13
HSU59209
LOCUS HSU59209 2107 bp mRNA linear PRI 02-JUL-1998
DEFINITION Homo sapiens C19steroid specific UDP-glucuronosyltransferase mRNA, complete cds.
ACCESSION U59209
VERSION U59209.1 GI:3287472
KEYWORDS UGT2B17G.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2107)
AUTHORS Beaulieu,M., Levesque,E., Hum,D.W. and Belanger,A.
TITLE Isolation and characterization of a novel cDNA encoding a human UDP-glucuronosyltransferase active on C19 steroids
JOURNAL J. Biol. Chem. 271 (37), 22855-22862 (1996)
MEDLINE 96394358
PUBMED 8798464
REFERENCE 2 (bases 1 to 2107)
AUTHORS Hum,D.W., Belanger,A., Beaulieu,M. and Levesque,E.
TITLE Direct Submission
JOURNAL Submitted (23-MAY-1996) Laboratory of Molecular Endocrinology, Centre Hospitalier de l'universite Laval, 2705 Boul. Laurier, Ste-Foy, Quebec G1V 4G2, Canada
FEATURES Location/Qualifiers
source 1. .2107
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="LNCaP"
/tissue_type="prostate"
CDS 52. .1644
/note="UGT2B17"
/codon_start=1
/product="C19steroid specific UDP-glucuronosyltransferase"
/protein_id="AAC25491.1"
/db_xref="GI:3287473"
/translation="MSLKWMSVFLLMQLSCYFSSGSCGKVLVWPTEYSHWINMKTILE
ELVQRGHEVIVLTSSASILVNASKSSAIKLEVYPTSLTKNDLEDFFMKMFDRWTYSIS
KNTFWSYFSQLQELCWEYSDYNIKLCEDAVLNKKLMRKLQESKFDVLLADAVNPCGEL
LAELLNIPFLYSLRFSVGYTVEKNGGGFLFPPSYVPVVMSELSDQMIFMERIKNMIYM
LYFDFWFQAYDLKKWDQFYSEVLGRPTTLFETMGKAEMWLIRTYWDFEFPRPFLPNVD
FVGGLHCKPAKPLPKEMEEFVQSSGENGIVVFSLGSMISNMSEESANMIASALAQIPQ
KVLWRFDGKKPNTLGSNTRLYKWLPQNDLLGHPKTKAFITHGGTNGIYEAIYHGIPMV
GIPLFADQHDNIAHMKAKGAALSVDIRTMSSRDLLNALKSVINDPIYKENIMKLSRIH
HDQPVKPLDRAVFWIEFVMRHKGAKHLRVAAHNLTWIQYHSLDVIAFLLACVATMIFM
ITKCCLFCFRKLAKTGKKKRD"
BASE COUNT 657 a 382 c 424 g 644 t
ORIGIN

Query Match 47.4%; Score 775.6; DB 9; Length 2107;
Best Local Similarity 68.6%; Pred. No. 6.8e-190;
Matches 1104; Conservative 0; Mismatches 494; Indels 12; Gaps 2;

```
Qy         24 ATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGCTCCTGCAGCTCTTCTGT---GTT 80

Db         46 ACCAGGATGTCTCTGAAATGGATGTCAGTCTTTCTGCTGATGCAGCTCAGTTGTTACTTT 105

Qy         81 GGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTGACATGAGCCATTGGCTTAAT 140

Db        106 AGCTCTGGGAGTTGTGGAAAGGTGCTGGTGTGGCCCACAGAATACAGCCATTGGATAAAT 165
```

```
Qy        141 GTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATGAGGTAACAGTATTGACTCAC 200
Db        166 ATGAAGACAATCCTGGAAGAGCTTGTTCAGAGGGGTCATGAGGTGATTGTGTTGACATCT 225

Qy        201 TCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCATTGAAATTTGAGGTGGTCCAT 260
Db        226 TCGGCTTCTATTCTTGTCAATGCCAGTAAATCATCTGCTATTAAATTAGAAGTTTATCCT 285

Qy        261 ATGCCACAGGACAGAACAGAAGAAAATGAAATATTT---------GTTGACCTAGCTCTG 311
Db        286 ACATCTTTAACTAAAAATGATTTGGAAGATTTTTTTATGAAAATGTTCGATAGATGGACA 345

Qy        312 AATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATGATTTTTTTGTT 371
Db        346 TATAGTATTTCAAAAAATACATTTTGGTCATATTTTTCACAACTACAAGAATTGTGTTGG 405

Qy        372 GAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATCAGACGCTTATG 431
Db        406 GAATATTCTGACTATAATATAAAGCTCTGTGAAGATGCAGTTTTGAACAAGAAACTTATG 465

Qy        432 AAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGATTCCCTGTGGA 491
Db        466 AGAAAACTACAAGAGTCAAAATTTGATGTCCTTCTGGCAGATGCCGTTAATCCCTGTGGT 525

Qy        492 GACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAATTTCTGTAGGA 551
Db        526 GAGCTGCTGGCTGAACTACTTAACATACCCTTTCTGTACAGTCTCCGCTTCTCTGTTGGC 585

Qy        552 GGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATGTACCTGTGCCT 611
Db        586 TACACAGTTGAGAAGAATGGTGGAGGATTTCTGTTCCCTCCTTCCTATGTACCTGTTGTT 645

Qy        612 ATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATTCAATGCTTTCA 671
Db        646 ATGTCAGAATTAAGTGATCAAATGATTTTCATGGAGAGGATAAAAAATATGATATATATG 705

Qy        672 GTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAGAGTTTTATAGT 731
Db        706 CTTTATTTTGACTTTTGGTTTCAAGCATATGATCTGAAGAAGTGGGACCAGTTTTATAGT 765

Qy        732 AAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTGAGATATGGCTA 791
Db        766 GAAGTTCTAGGAAGACCCACTACATTATTTGAGACAATGGGGAAAGCTGAAATGTGGCTC 825

Qy        792 ATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACTTTGAGTTTGTT 851
Db        826 ATTCGAACCTATTGGGATTTTGAATTTCCTCGCCCATTCTTACCAAATGTTGATTTTGTT 885

Qy        852 GGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAAATTTTGTCCAG 911
Db        886 GGAGGACTTCACTGTAAACCAGCCAAACCCTTGCCTAAGGAAATGGAAGAGTTTGTGCAG 945

Qy        912 AGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTCAAAATGTTACA 971
Db        946 AGCTCTGGAGAAAATGGTATTGTGGTGTTTTCTCTGGGGTCGATGATCAGTAACATGTCA 1005

Qy        972 GAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGAAGGTGTTATGG 1031
Db       1006 GAAGAAAGTGCCAACATGATTGCATCAGCCCTTGCCCAGATCCCACAAAAGGTTCTATGG 1065

Qy       1032 AGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGTATGATTGGATA 1091
Db       1066 AGATTTGATGGCAAGAAGCCAAATACTTTAGGTTCCAATACTCGACTGTATAAGTGGTTA 1125

Qy       1092 CCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTCATGGTGGAATG 1151
Db       1126 CCCCAGAATGACCTTCTTGGTCATCCCAAAACCAAAGCTTTTATAACTCATGGTGGAACC 1185

Qy       1152 AATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTCCCATATTTGGT 1211
Db       1186 AATGGCATCTATGAGGCGATCTACCATGGGATCCCTATGGTGGGCATTCCCTTGTTTGCG 1245

Qy       1212 GATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAGAAATAAACTTC 1271

Db       1246 GATCAACATGATAACATTGCTCACATGAAAGCCAAGGGAGCAGCCCTCAGTGTGGACATC 1305

Qy       1272 AAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTACCGATTCCTCT 1331
Db       1306 AGGACCATGTCAAGTAGAGATTTGCTCAATGCATTGAAGTCAGTCATTAATGACCCTATC 1365

Qy       1332 TATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTGTAAAGCCCCTA 1391
Db       1366 TATAAAGAGAATATCATGAAATTATCAAGAATTCATCATGATCAACCGGTGAAGCCCCTG 1425

Qy       1392 GATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCAAGCACCTGCGA 1451
Db       1426 GATCGAGCAGTCTTCTGGATTGAGTTTGTCATGCGCCATAAAGGAGCCAAGCACCTTCGG 1485

Qy       1452 TCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGATTGGGTTCCTG 1511
Db       1486 GTCGCAGCCCACAACCTCACCTGGATCCAGTACCACTCTTTGGATGTGATAGCATTCCTG 1545

Qy       1512 CTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTATTTTCCTGTCAA 1571
Db       1546 CTGGCCTGCGTGGCAACTATGATATTTATGATCACAAAATGTTGCCTGTTTTGTTTCCGA 1605

Qy       1572 AAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAAA 1621
Db       1606 AAGCTTGCCAAAACAGGAAAGAAGAAGAAAAGGGATTAGTTATATCAAAA 1655
```

```
RESULT 14
AF112112
LOCUS       AF112112               2108 bp    mRNA    linear   PRI 11-APR-1999
DEFINITION  Macaca fascicularis UDP-glucuronosyltransferase 2B19 precursor,
            mRNA, complete cds.
ACCESSION   AF112112
VERSION     AF112112.1  GI:4580601
KEYWORDS    .
SOURCE      Macaca fascicularis.
  ORGANISM  Macaca fascicularis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Macaca.
REFERENCE   1  (bases 1 to 2108)
  AUTHORS   Belanger,G., Barbier,O., Hum,D.W. and Belanger,A.
  TITLE     Molecular cloning, expression and characterization of a monkey
            steroid UDP-glucuronosyltransferase, UGT2B19, that conjugates
            testosterone
  JOURNAL   Eur. J. Biochem. 260 (3), 701-708 (1999)
  MEDLINE   99203465
   PUBMED   10102998
REFERENCE   2  (bases 1 to 2108)
  AUTHORS   Belanger,G., Barbier,O., Hum,D.W. and Belanger,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-DEC-1998) Molecular Endocrinology, CHUL Research
            Center, 2705 Laurier Blvd, Sainte Foy, Quebec G1V 4G2, Canada
FEATURES             Location/Qualifiers
     source          1. .2108
                     /organism="Macaca fascicularis"
                     /db_xref="taxon:9541"
     CDS             26. .1612
                     /EC_number="2.4.1.17"
                     /note="UGT2B19"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase 2B19 precursor"
                     /protein_id="AAD24435.1"
                     /db_xref="GI:4580602"
                     /translation="MSMKWTSALLLIQLSCYLSFGSCGKVLVWPTEFSHWMNIKTILD
                     ELVQRGHEVTVLAYSTSILPDPNNPSPLKFEICPTSLTETEFQDSVTQLVKRWSDIRK
                     DTFWPHFLHVQEMMWTYGDMIRKFCKDVVSNKKLMKKLQESRFDVVLADAISPCGELL
                     AELLKIPFVYSLRFSPGYALEKHGGGFLFPPSYVPVTMSELRDQMTFMERVQNMIYMV
                     YFDFWFQVWDVKNWDQFYSKVLGRPTTLFEIMAKAEIWLIRNYWDFQFPHPLLPNVEF
                     VGGLHCKPAKPLPKEMEEFVQSSGDNGVVVFSLGSMVSNMSEERANVIASALAKIPQK
                     VLWRFDGNKPDTLGLNTQLYKWLPQNDLLGHPKTRAFITHGGANGIYEAIYHGIPMVG
                     VPLFADQPDNIAHMKAKGAAVRLDFDTMSSTDLLNALKTVINDPIYKENAMKLSSIHH
```

DQPVKPLDRAVFWIEFVMRHKGAKHLRVAAHDLTWFQYHSLDVIGFLLACVATVIFII
TKCLFCVWKFVRTRKKGKRD"
BASE COUNT 649 a 393 c 438 g 628 t
ORIGIN

Query Match 46.9%; Score 767.6; DB 9; Length 2108;
Best Local Similarity 68.8%; Pred. No. 8.1e-188;
Matches 1103; Conservative 0; Mismatches 489; Indels 12; Gaps 3;

```
Qy         23 CATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGCTCCTGCAGCTCTTCTGT---GT  79
Db         19 CATCAGGATGTCTATGAAATGGACTTCAGCTCTTCTGCTGATACAGCTGAGCTGTTACTT  78

Qy         80 TGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTGACATGAGCCATTGGCTTAA 139
Db         79 GAGCTTTGGGAGTTGTGGAAAGGTGCTGGTGTGGCCCACAGAATTCAGCCATTGGATGAA 138

Qy        140 TGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATGAGGTAACAGTATTGACTCA 199
Db        139 TATAAAGACAATTCTGGATGAGCTTGTCCAGAGAGGTCATGAGGTGACTGTACTGGCATA 198

Qy        200 CTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCATTGAAATTTGAGGTGGTCCA 259
Db        199 TTCAACTTCCATTCTTCCTGATCCCAACAACCCATCTCCTCTTAAATTTGAAATTTGTCC 258

Qy        260 TATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTGACCTAGCTCTGA------A 313
Db        259 TACATCTTTAACCGAAACTGAGTTTCAGGATAGCGTCACGCAACTGGTTAAGAGATGGTC 318

Qy        314 TGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATGATTTTTTTGTTGA 373
Db        319 AGACATTCGAAAAGACACATTTTGGCCACATTTTTTACACGTACAAGAAATGATGTGGAC 378

Qy        374 AATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATCAGACGCTTATGAA 433
Db        379 ATATGGTGACATGATTAGAAAGTTCTGTAAAGATGTGGTTTCTAATAAGAAACTTATGAA 438

Qy        434 GAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGATTCCCTGTGGAGA 493
Db        439 GAAACTACAGGAGTCAAGATTTGATGTTGTTCTTGCGGATGCTATTTCCCCCTGTGGTGA 498

Qy        494 CCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAATTTCTGTAGGAGG 553
Db        499 GCTACTGGCTGAGCTACTTAAAATACCCTTTGTGTACAGCCTCCGCTTCTCTCCTGGCTA 558

Qy        554 CAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATGTACCTGTGCCTAT 613
Db        559 CGCACTTGAAAAGCATGGTGGAGGGTTTCTGTTCCCTCCTTCCTATGTACCTGTTACTAT 618

Qy        614 GACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATTCAATGCTTTCAGT 673
Db        619 GTCAGAATTAAGAGATCAAATGACTTTCATGGAGAGAGTACAAAATATGATCTATATGGT 678

Qy        674 TTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAGAGTTTTATAGTAA 733
Db        679 TTATTTTGACTTTTGGTTCCAAGTATGGGATGTAAAGAACTGGGATCAGTTTTACAGTAA 738

Qy        734 GGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTGAGATATGGCTAAT 793
Db        739 AGTTCTAGGAAGACCCACTACATTATTTGAGATAATGGCAAAAGCTGAAATATGGCTTAT 798

Qy        794 ACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACTTTGAGTTTGTTGG 853
Db        799 TCGAAACTACTGGGATTTTCAATTTCCTCATCCACTCTTACCAAATGTTGAGTTCGTTGG 858

Qy        854 AGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAAATTTTGTCCAGAG 913
Db        859 AGGACTCCACTGCAAACCTGCCAAACCCCTACCTAAGGAAATGGAAGAGTTCGTCCAGAG 918

Qy        914 TTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTCAAAATGTTACAGA 973
Db        919 CTCTGGAGACAATGGTGTTGTGGTGTTTTCTCTGGGGTCCATGGTCAGTAACATGTCAGA 978

Qy        974 AGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGAAGGTGTTATGGAG 1033
Db        979 AGAAAGGGCCAACGTAATTGCATCAGCCCTTGCCAAGATCCCACAAAAGGTTCTGTGGAG 1038

Qy       1034 GTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGTATGATTGGATACC 1093
Db       1039 ATTTGATGGGAATAAACCAGATACCTTAGGACTCAATACTCAGCTGTACAAGTGGCTACC 1098

Qy       1094 CCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTCATGGTGGAATGAA 1153
Db       1099 CCAGAATGATCTTCTTGGTCATCCAAAAACCAGAGCTTTTATAACTCACGGTGGAGCCAA 1158

Qy       1154 TGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTCCCATATTTGGTGA 1213
Db       1159 TGGCATCTACGAGGCAATCTACCATGGAATCCCTATGGTGGGCGTTCCATTGTTTGCCGA 1218

Qy       1214 TCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAGAAATAAACTTCAA 1273
Db       1219 TCAACCTGATAACATTGCTCACATGAAGGCCAAAGGAGCAGCTGTTAGACTGGACTTCGA 1278

Qy       1274 AACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTACCGATTCCTCTTA 1333
Db       1279 CACAATGTCGAGTACAGACTTACTCAATGCACTGAAGACAGTAATAAATGATCCTATATA 1338

Qy       1334 TAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTGTAAAGCCCCTAGA 1393
Db       1339 TAAAGAGAATGCTATGAAATTATCAAGCATTCATCATGATCAACCAGTGAAGCCCCTGGA 1398

Qy       1394 TCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCAAGCACCTGCGATC 1453
Db       1399 TCGAGCAGTCTTCTGGATTGAATTTGTCATGCGCCATAAAGGAGCCAAACACCTTCGGGT 1458

Qy       1454 AGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGATTGGGTTCCTGCT 1513
Db       1459 TGCAGCCCATGACCTCACCTGGTTCCAGTACCACTCTTTGGATGTGATTGGGTTCCTGCT 1518

Qy       1514 GACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTATTTTCCTGTCAAAA 1573
Db       1519 GGCCTGTGTGGCAACTGTGATATTCATCATCACAAAATGTCTGTTTTGTGTCTG---GAA 1575

Qy       1574 ATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTC 1617
Db       1576 GTTTGTTAGAACAAGAAAGAAGGGAAAAAGAGATTAAATACATC 1619
```

RESULT 15
AX410640
LOCUS AX410640 2123 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 3287 from Patent WO0229103.
ACCESSION AX410640
VERSION AX410640.1 GI:21443345
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Alvares,C., Horne,D., Peres-da-Silva,S. and Vockley,J.G.
TITLE Gene expression profiles in liver cancer
JOURNAL Patent: WO 0229103-A 3287 11-APR-2002;
GENE LOGIC INC (US)
FEATURES Location/Qualifiers
source 1. .2123
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="EMBL/GenBank Accession No. U06641"
BASE COUNT 676 a 389 c 427 g 631 t
ORIGIN

Query Match 46.8%; Score 766.2; DB 6; Length 2123;
Best Local Similarity 68.4%; Pred. No. 1.9e-187;
Matches 1078; Conservative 0; Mismatches 489; Indels 9; Gaps 1;

```
Qy        55 TTCTGCTCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGC 114
Db         8 TGCTGATACAGCTCAGTTGTTACTTTAGCTCTGGAAGCTGTGGAAAGGTGCTAGTGTGGC  67

Qy       115 CCTGTGACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAG 174
Db        68 CCACAGAATACAGCCATTGGATAAATATGAAGACAATCCTGGAAGAGCTTGTTCAGAGGG 127

Qy       175 GCCATGAGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTT 234
Db       128 GTCATGAGGTGACTGTGTTGACATCTTCGGCTTCTACTCTTGTCAATGCCAGTAAATCAT 187

Qy       235 CTGCATTGAAATTTGAGGTGGTCCATATGCC---------ACAGGACAGAACAGAAGAAA 285
Db       188 CTGCTATTAAATTAGAAGTTTATCCTACATCTTTAACTAAAAATTATTTGGAAGATTCTC 247

Qy       286 ATGAAATATTTGTTGACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAG 345
Db       248 TTCTGAAAATTCTCGATAGATGGATATATGGTGTTTCAAAAAATACATTTTGGTCATATT 307

Qy       346 TTATAAAATTAAATGATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGA 405
Db       308 TTTCACAATTACAAGAATTGTGTTGGGCATATTATGACTACAGTAACAAGCTCTGTAAAG 367

Qy       406 GCTTTATCTACAATCAGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGC 465
Db       368 ATGCAGTTTTGAATAAGAAACTTATGATGAAACTACAAGAGTCAAGATTTGATGTCATTC 427

Qy       466 TTATAGACCCTGTGATTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTG 525
Db       428 CTGGAGATCCTGTTTTCCCCTGTGGTGAGCTACTGTCAGAGCTACTTAACATACCCTTTG 487

Qy       526 TGCTCACACTTAGAATTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAG 585
Db       488 TCTACAGGTCTCGCATCTCTCGTGGCTACACAATTGAGAAGAATGGTGGAGGATTTCTGT 547

Qy       586 CTCCACTTTCCTATGTACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGG 645
Db       548 TCCCTCCTTCCTATGTACCTGTTGTTATGTCAGAATTAATTGATCAAATGATTTTCATGG 607

Qy       646 AAAGAGTAAAAAATTCAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACT 705
Db       608 AGAGGATAAAAAATATGATACATATGCTTTATTTTGACTTTTGGTTTCAAATTTATGATC 667

Qy       706 ATCATTTTTGGGAAGAGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGA 765
Db       668 TGAAGAAGTGGGACCAGTTTTATAGTGAAGTTCTAGGAAGACCCACTACATTATTTGAGA 727

Qy       766 CTGTGGGAAAAGCTGAGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAAC 825
Db       728 CAATGGGGAAAGCTGAAATGTGGCTCATTCGAACCTATTGGGATTTTGAATTTCCTCGCC 787

Qy       826 CATACCAACCTAACTTTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGC 885
Db       788 CATTCTTACCAAATGTTGATTTTGTTGGAGGACTTCACTGTAAACCAGCCAAACCCCTGC 847

Qy       886 CTAAGGAAATGGAAAATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTC 945
Db       848 CTAAGGAATTGGAAGAGTTTGTGCAGAGCTCTGGAGAAAATGGTATTGTGGTGTTTTCTC 907

Qy       946 TGGGGTCACTGTTTCAAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTG 1005
Db       908 TGGGGTCGATGATCAGTAACATGTCAGAAGAAAGTGCCAACATGATTGCATCAGCCCTTG 967

Qy      1006 CCCAGATCCCACAGAAGGTGTTATGGAGGTACAAAGGAAAAAAACCATCCACATTAGGAG 1065
Db       968 CCCAGATCCCACAAAAGGTTCTATGGAGATTTGATGGCAAGAAGCCAAATACTTTAGGTT 1027

Qy      1066 CCAATACTCGGCTGTATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCA 1125
Db      1028 CCAATACTCGACTGTACAAGTGGTTACCCCAGAATGACCTTCTTGGTCATCCCAAAACCA 1087

Qy      1126 AAGCTTTTATCACTCATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCC 1185

Db      1088 AAGCTTTTATAACTCATGGTGGAACCAATGGCATCTATGAGGCGATCTACCATGGGATCC 1147

Qy      1186 CTATGGTGGGAGTTCCCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCA 1245
Db      1148 CTATGGTGGGCATTCCCTTGTTTGCGGATCAAGATGATAACATTGCTCACATGAAAGCCA 1207

Qy      1246 AAGGAGCAGCTGTAGAAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTT 1305
Db      1208 AGGGAGCAGCCCTCAGTGTGGACATCAGGACCATGTCAAGTAGAGATTTGCTCAATGCAT 1267

Qy      1306 TGAGAACAGTCATTACCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTC 1365
Db      1268 TGAAGTCAGTCATTAATGACCCAGTCTATAAAGAGAATGCTATGAAATTATCAAGAATTC 1327

Qy      1366 ACCATGATCAACCTGTAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGC 1425
Db      1328 ATCATGACCAACCAATGAAGCCCCTGGATCGAGCAGTCTTCTGGATTGAGTTTGTCATGC 1387

Qy      1426 GCCACAAAGGAGCCAAGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACT 1485
Db      1388 GCCACAAAGGAGCCAAGCACCTTCGAGTCGCAGCTCACAACCTCACCTGGATCCAGTACC 1447

Qy      1486 ACTCTATAGATGTGATTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCA 1545
Db      1448 ACTCTTTGGATGTGATAGCATTCCTGCTGGCCTGGGTGGCAACTGTGATATTTATCATCA 1507

Qy      1546 CAAAATGTTTTTTATTTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGG 1605
Db      1508 CAAAATTTTGCCTGTTTTGTTTCCGAAAGCTTGCCAAAAAAGGAAAGAAGAAGAAAAGAG 1567

Qy      1606 AATAGATCTTTCCAAA 1621
Db      1568 ATTAGTTATATCAAAA 1583
```

Search completed: July 16, 2003, 13:28:03
Job time : 2983.96 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 16, 2003, 03:41:41 ; Search time 279.458 Seconds
(without alignments)
13183.634 Million cell updates/sec

Title: US-09-981-353-165
Perfect score: 1636
Sequence: 1 gatcagtgtgtgagggaact.....ccaaattcaagaaagacctg 1636

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1636	100.0	1636	24	AAD24666	Human drug metabol
2	1636	100.0	2759	24	ABK12422	cDNA encoding huma
3	1636	100.0	2974	22	AAS46185	Human DNA encoding
4	967.4	59.1	1650	21	AAC65396	Human carbohydrate
5	821.2	50.2	1766	23	AAS91380	DNA encoding novel
6	782	47.8	2090	24	ABK84210	Human cDNA differe
7	782	47.8	2090	24	ABN96795	Gene #3293 used to
8	780.4	47.7	1976	21	AAZ95206	Human UDP-glucuron
9	775.6	47.4	2107	19	AAV15900	Uridine diphospho-
10	766.2	46.8	2123	24	ABN96789	Gene #3287 used to
11	766	46.8	2092	21	AAZ95199	Human UDP-glucuron
12	758	46.3	2093	24	ABN97347	Gene #3845 used to
13	754.8	46.1	1722	22	AAD06860	Human breast cance
14	754.8	46.1	1855	24	ABN95622	Gene #2120 used to
15	754.8	46.1	1855	24	ABL68501	Kidney cancer rela
16	754.8	46.1	1855	24	ABL68868	Kidney cancer rela
17	753.4	46.1	2111	23	AAS69712	DNA encoding novel
18	750	45.8	1854	21	AAZ95200	Human UDP-glucuron
19	742.8	45.4	1713	22	AAD06844	Human breast cance
20	732.6	44.8	2799	24	ABN97261	Gene #3759 used to
21	732.2	44.8	21000	24	ABK12423	cDNA encoding huma
22	717.6	43.9	1859	23	AAS69710	DNA encoding novel
23	714.8	43.7	2802	23	AAS69711	DNA encoding novel
24	693.4	42.4	1961	24	ABK63496	Rat sequence diffe
25	486.4	29.7	515	20	AAV87412	EST clone BR77. H
26	350	21.4	350	21	AAC03286	Human secreted pro
27	341.2	20.9	2406	23	AAS91382	DNA encoding novel
28	329	20.1	2422	24	ABN95608	Gene #2106 used to
29	329	20.1	2422	24	ABL68883	Kidney cancer rela
30	327.2	20.0	1655	23	AAS91252	DNA encoding novel
31	322.4	19.7	2320	24	ABA99240	Human UGT1A9 encod
32	322.4	19.7	2320	24	ABA99241	Human UGT1A9 encod
33	322.4	19.7	2320	24	AAD30177	Human UDP-glucuron
34	322.4	19.7	2320	24	ABA05198	Human uridine 5'di
35	322.4	19.7	2320	24	ABA05199	Human uridine 5'di
c 36	319.2	19.5	1001	21	AAH51497	Human UGT2B7 relat
37	317.8	19.4	2368	13	AAQ27370	HUG-Br2. Homo sap
38	316.6	19.4	2351	13	AAQ27369	HUG-Br1. Homo sap
c 39	314.2	19.2	1001	21	AAH51498	Human UGT2B7 relat
40	272.2	16.6	735	21	AAZ45118	UDP-glucuronosyltr
41	268	16.4	556	24	ABN65577	Human cancer relat
42	256.8	15.7	874	24	ABL01468	Murine apoptosis r
43	250.8	15.3	2308	24	ABI99695	Mouse ischaemic co
c 44	248	15.2	2448	23	AAS72022	DNA encoding novel
c 45	248	15.2	2448	23	AAS74393	DNA encoding novel

ALIGNMENTS

RESULT 1
AAD24666
ID AAD24666 standard; cDNA; 1636 BP.
XX
AC AAD24666;
XX
DT 12-MAR-2002 (first entry)
XX
DE Human drug metabolising enzyme (DME)-1 cDNA.
XX
KW Human; drug metabolising enzyme; gene therapy; autoimmune disorder;
KW inflammatory disorder; acquired immune deficiency syndrome; infection;
KW AIDS; adult respiratory distress syndrome; cell proliferative disorder;
KW allergy; anaemia; conjunctivitis; actinic keratosis; arteriosclerosis;
KW cancer; endocrine disorder; hypothalamus disorder; pituitary disorder;
KW gastrointestinal disorder; metabolic disorder; developmental disorder;
KW liver disorder; iritis; cystic fibrosis; Addison's disease; retinitis;
KW goitre; diabetes; dysphagia; gastric carcinoma; anorexia; eye disorder;
KW DME-1; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 27..1610
FT /*tag= a
FT /product= "Human DME-1 protein"
FT sig_peptide 27..95
FT /*tag= b
FT mat_peptide 96..1607
FT /*tag= c
FT /product= "Human mature DME-1 protein"
XX

PN WO200179468-A2.
XX
PD 25-OCT-2001.
XX
PF 12-APR-2001; 2001WO-US11869.
XX
PR 13-APR-2000; 2000US-197590P.
PR 19-APR-2000; 2000US-198403P.
PR 28-APR-2000; 2000US-200185P.
PR 05-MAY-2000; 2000US-202234P.
PR 11-MAY-2000; 2000US-203509P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Policky JL, Hafalia A, Burford N, Ring HZ, Lal P, Tribouley CM;
PI Yao MG, Yue H, Tang YT, Patterson C, Das D, Sanjanwala MS;
PI Gandhi AR, Reddy R, Khan FA, Baughn MR, Ramkumar J, Griffin JA;
PI Au-Young J;
XX
DR WPI; 2002-066363/09.
DR P-PSDB; AAE15434.
XX
PT Novel isolated human drug metabolizing enzymes referred as DME 1-10,
PT useful for diagnosing, treating, or preventing disorders associated
PT with aberrant expression of DME such as allergy, anemia, asthma,
PT infertility -
XX
PS Claim 5; Page 136-137; 143pp; English.
XX
CC The invention relates to human drug metabolising enzymes referred as
CC DME and nucleic acid molecules encoding such enzymes. Polynucleotides
CC of the invention are useful for assessing toxicity of test compounds
CC and in gene therapy. Sequences of the invention are useful in the
CC diagnosis, prevention and treatment of autoimmune/inflammatory
CC disorders such as acquired immune deficiency syndrome (AIDS), adult
CC respiratory distress syndrome, allergies, anaemia, atherosclerosis,
CC asthma, autoimmune haemolytic anaemia, contact dermatitis, Crohn's
CC disease, glomerulonephritis, Goodpasture's syndrome, gout, Grave's
CC disease, Hashimoto's thyroiditis, irritable bowel syndrome, multiple
CC sclerosis, osteoarthritis, osteoporosis, psoriasis, systemic lupus
CC erythematosus, rheumatoid arthritis, scleroderma, ulcerative colitis,
CC uveitis, viral, bacterial, fungal, parasitic, protozoal, helminthic
CC infections; cell proliferative disorders such as actinic keratosis,
CC arteriosclerosis, atherosclerosis, Duchenne and Becker dystrophy,
CC hepatitis, Cushing's syndrome, cancers, myelodysplastic syndrome,
CC epilepsy; endocrine disorders such as disorders of the hypothalamus
CC and pituitary resulting from lesions such as primary brain tumours,
CC adenomas, infarction associated with pregnancy, aneurysms, vascular
CC malformations; eye disorders such as conjunctivitis, iritis, retinitis,
CC glaucoma, pigmentosa; metabolic disorders such as Addison's disease,
CC cystic fibrosis, diabetes, goitre, glycogen storage diseases,
CC hypercholesterolaemia, hyperthyroidism, hypoglycaemia, lipid myopathies,
CC Menkes syndrome, mannosidosis, obesity; gastrointestinal disorders such
CC as dysphagia, gastric carcinoma, anorexia, nausea, gastroenteritis,
CC hyperbilirubinaemia, emesis, cirrhosis, diarrhoea, jaundice, Reye's
CC syndrome, peliosis hepatitis, hepatic vein thrombosis and developmental
CC disorders. The present sequence is human DME-1 protein cDNA.
XX
SQ Sequence 1636 BP; 483 A; 315 C; 369 G; 469 T; 0 other;

Query Match 100.0%; Score 1636; DB 24; Length 1636;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1636; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GATCAGTGTGTGAGGGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GATCAGTGTGTGAGGGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGC 60

Qy         61 TCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTG 120

Qy        121 ACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATG 180

Qy        181 AGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCAT 240

Qy        241 TGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTG 300

Qy        301 ACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATG 360

Qy        361 ATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATC 420

Qy        421 AGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGA 480

Qy        481 TTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAA 540

Qy        541 TTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATG 600

Qy        601 TACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATT 660

Qy        661 CAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAG 720

Qy        721 AGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTG 780

Qy        781 AGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACT 840

Qy        841 TTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAA 900

Qy        901 ATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTC 960

Qy        961 AAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 AAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGA 1020

Qy       1021 AGGTGTTATGGAGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGGTGTTATGGAGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGT 1080

Qy       1081 ATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTC 1140

Qy       1141 ATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTC 1200

Qy       1201 CCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1201 CCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAG 1260

Qy       1261 AAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTA 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTA 1320

Qy       1321 CCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTG 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTG 1380

Qy       1381 TAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCA 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCA 1440

Qy       1441 AGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGA 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGA 1500

Qy       1501 TTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTAT 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTAT 1560

Qy       1561 TTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1620

Qy       1621 ATTCAAGAAAGACCTG 1636
             ||||||||||||||||
Db       1621 ATTCAAGAAAGACCTG 1636
```

```
RESULT 2
ABK12422
ID   ABK12422 standard; cDNA; 2759 BP.
XX
AC   ABK12422;
XX
DT   18-JUN-2002  (first entry)
XX
DE   cDNA encoding human drug-metabolising enzyme.
XX
KW   Human; drug-metabolising enzyme; chromosome 4; therapeutic agent;
KW   UDP-glucuronosyltransferase drug-metabolising enzyme;
KW   drug screening assay; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           1..37
FT                   /*tag= a
FT   CDS             38..1621
FT                   /*tag= b
FT                   /product= "Drug-metabolising enzyme"
FT   3'UTR           1622..2759
FT                   /*tag= c
XX
PN   WO200218554-A2.
XX
PD   07-MAR-2002.
XX
PF   24-AUG-2001; 2001WO-US26403.
XX
PR   30-AUG-2000; 2000US-228893P.
PR   16-FEB-2001; 2001US-0784340.
XX
PA   (APPL-) APPLERA CORP.
XX
PI   Wei M,  Ye J,  Ketchum KA,  Di Francesco V,  Beasley EM;
XX
DR   WPI; 2002-304252/34.
DR   P-PSDB; AAU77927.
XX
PT   A human drug-metabolising enzyme, related to the
PT   UDP-glucuronosyltransferase drug-metabolising enzyme subfamily, used as
PT   model for developing human therapeutic targets and as a target for
PT   human therapeutics -
XX
PS   Claim 4; Fig 1; 86pp; English.
XX
CC   The present invention relates to the isolation of a human
CC   drug-metabolising enzyme, and polynucleotide sequences encoding it.
CC   The gene encoding human drug-metabolising enzyme maps to chromosome 4.
CC   The drug-metabolising enzyme is related to the
CC   UDP-glucuronosyltransferase drug-metabolising enzyme subfamily, The
CC   drug-metabolising enzyme, its allelic variant, orthologue or fragment,
CC   are useful for identifying an agent that binds or modulates it. A
CC   pharmaceutical composition comprising an agent that binds the enzyme is
CC   useful for treating a disease or condition mediated by human
CC   drug-metabolising enzyme. The enzyme can be used as a model for
CC   the development of human therapeutic targets, aid in the identification
CC   of therapeutic proteins and serve as targets for the development of
CC   human therapeutic agents that modulate drug-metabolising enzyme
CC   activity in cells and tissues that express the drug-metabolising enzyme.
CC   The enzyme can also be used in screening assays to screen a compound for
CC   its ability to stimulate or inhibit interaction between a
CC   drug-metabolising enzyme and a molecule that normally interacts with the
CC   drug-metabolising enzyme. The nucleic acid molecules are useful for
CC   constructing recombinant vectors, host cells and transgenic animals,
CC   and for designing ribozymes. The nucleic acids are also useful in drug
CC   screening assays and as a target for treatment by the compounds
CC   identified through drug screening. The present sequence encodes for
CC   human drug-metabolising enzyme.
XX
SQ   Sequence 2759 BP; 846 A; 520 C; 544 G; 849 T; 0 other;
```

```
  Query Match             100.0%;  Score 1636;  DB 24;  Length 2759;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1636;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GATCAGTGTGTGAGGGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         12 GATCAGTGTGTGAGGGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGC 71

Qy         61 TCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         72 TCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTG 131

Qy        121 ACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        132 ACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATG 191

Qy        181 AGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        192 AGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCAT 251

Qy        241 TGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        252 TGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTG 311

Qy        301 ACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        312 ACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATG 371

Qy        361 ATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        372 ATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATC 431

Qy        421 AGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        432 AGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGA 491

Qy        481 TTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        492 TTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAA 551

Qy        541 TTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATG 600
```

```
Db        552 TTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATG 611

Qy        601 TACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        612 TACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATT 671

Qy        661 CAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        672 CAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAG 731

Qy        721 AGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        732 AGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTG 791

Qy        781 AGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        792 AGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACT 851

Qy        841 TTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        852 TTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAA 911

Qy        901 ATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        912 ATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTC 971

Qy        961 AAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        972 AAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGA 1031

Qy       1021 AGGTGTTATGGAGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1032 AGGTGTTATGGAGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGT 1091

Qy       1081 ATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1092 ATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTC 1151

Qy       1141 ATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1152 ATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTC 1211

Qy       1201 CCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1212 CCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAG 1271

Qy       1261 AAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1272 AAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTA 1331

Qy       1321 CCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1332 CCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTG 1391

Qy       1381 TAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1392 TAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCA 1451

Qy       1441 AGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1452 AGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGA 1511

Qy       1501 TTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTAT 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1512 TTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTAT 1571

Qy       1561 TTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1572 TTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1631

Qy       1621 ATTCAAGAAAGACCTG 1636
              ||||||||||||||||
Db       1632 ATTCAAGAAAGACCTG 1647
```

RESULT 3
AAS46185
ID AAS46185 standard; cDNA; 2974 BP.
XX
AC AAS46185;
XX
DT 18-DEC-2001 (first entry)
XX
DE Human DNA encoding PRO polypeptide sequence #261.
XX
KW PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep; ss;
KW dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
KW blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
KW adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder;
KW PCR primer.
XX
OS Homo sapiens.
XX
PN WO200168848-A2.
XX
PD 20-SEP-2001.
XX
PF 28-FEB-2001; 2001WO-US06520.
XX
PR 01-MAR-2000; 2000WO-US05601.
PR 02-MAR-2000; 2000WO-US05841.
PR 03-MAR-2000; 2000US-187202P.
PR 06-MAR-2000; 2000US-186968P.
PR 14-MAR-2000; 2000US-189320P.
PR 14-MAR-2000; 2000US-189328P.
PR 15-MAR-2000; 2000WO-US06884.
PR 21-MAR-2000; 2000US-190828P.
PR 21-MAR-2000; 2000US-191007P.
PR 21-MAR-2000; 2000US-191048P.
PR 21-MAR-2000; 2000US-191314P.
PR 28-MAR-2000; 2000US-192655P.
PR 29-MAR-2000; 2000US-193032P.
PR 29-MAR-2000; 2000US-193053P.
PR 30-MAR-2000; 2000WO-US08439.
PR 04-APR-2000; 2000US-194449P.
PR 04-APR-2000; 2000US-194647P.
PR 11-APR-2000; 2000US-195975P.
PR 11-APR-2000; 2000US-196000P.
PR 11-APR-2000; 2000US-196187P.
PR 11-APR-2000; 2000US-196690P.
PR 11-APR-2000; 2000US-196820P.
PR 18-APR-2000; 2000US-198121P.
PR 18-APR-2000; 2000US-198585P.
PR 25-APR-2000; 2000US-199397P.
PR 25-APR-2000; 2000US-199550P.
PR 25-APR-2000; 2000US-199654P.
PR 03-MAY-2000; 2000US-201516P.
PR 17-MAY-2000; 2000WO-US13705.
PR 22-MAY-2000; 2000WO-US14042.
PR 30-MAY-2000; 2000WO-US14941.
PR 02-JUN-2000; 2000WO-US15264.
PR 05-JUN-2000; 2000US-209832P.
PR 28-JUL-2000; 2000WO-US20710.
PR 22-AUG-2000; 2000US-0644848.
PR 24-AUG-2000; 2000WO-US23328.
PR 08-NOV-2000; 2000WO-US30952.
PR 01-DEC-2000; 2000WO-US32678.
PR 20-DEC-2000; 2000WO-US34956.
XX
PA (GETH) GENENTECH INC.
XX
PI Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;
PI Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;
XX
DR WPI; 2001-602746/68.

DR P-PSDB; AAU29284.
XX
PT Novel nucleic acids encoding PRO polypeptides, used to diagnose the
PT presence of tumours, such as prostate and breast tumours, in mammals and
PT to screen for modulators of the compounds -
XX
PS Claim 2; Fig 521; 774pp; English.
XX
CC Sequences AAS45925-AAS46231 represent DNA molecules encoding and PCR
CC primers for PRO polypeptides of the invention. The sequences of the
CC invention can be used to detect the presence of a tumour in a mammal by
CC comparing the level of expression of a PRO polypeptide in a test sample
CC of cells from the animal and a control sample of normal cells, whereby a
CC higher level of expression in the test sample indicates the presence of a
CC tumour in the mammal. Mammals include dogs, cats, cattle, horses, sheep,
CC pigs, goats and rabbits but are preferably human. The polypeptides can be
CC used to stimulate tumour necrosis factor (TNF) alpha release from human
CC blood, when contacted with it. A specific polypeptide can be used to
CC stimulate the proliferation or differentiation of chondrocyte cells. The
CC PRO proteins can be used to determine the presence of tumours and also
CC susceptibility to tumour development, particularly adrenal, lung, colon,
CC breast, prostate, rectal, cervical, or liver tumours, in mammalian
CC subjects. The oligonucleotide probes specific for the PRO nucleic acids
CC can be used for genetic analysis of individuals with genetic disorders.
XX
SQ Sequence 2974 BP; 946 A; 497 C; 556 G; 975 T; 0 other;

Query Match 100.0%; Score 1636; DB 22; Length 2974;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1636; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GATCAGTGTGTGAGGGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGC  60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GATCAGTGTGTGAGGGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGC  60

Qy         61 TCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTG 120

Qy        121 ACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATG 180

Qy        181 AGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCAT 240

Qy        241 TGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTG 300

Qy        301 ACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATG 360

Qy        361 ATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATC 420

Qy        421 AGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGA 480

Qy        481 TTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAA 540

Qy        541 TTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATG 600

Qy        601 TACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATT 660

Qy        661 CAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAG 720

Qy        721 AGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTG 780

Qy        781 AGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACT 840

Qy        841 TTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAA 900

Qy        901 ATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTC 960

Qy        961 AAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 AAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGA 1020

Qy       1021 AGGTGTTATGGAGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGGTGTTATGGAGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGT 1080

Qy       1081 ATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTC 1140

Qy       1141 ATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTC 1200

Qy       1201 CCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAG 1260

Qy       1261 AAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTA 1320

Qy       1321 CCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTG 1380

Qy       1381 TAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCA 1440

Qy       1441 AGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGA 1500

Qy       1501 TTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTAT 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTAT 1560

Qy       1561 TTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1620

Qy       1621 ATTCAAGAAAGACCTG 1636
              ||||||||||||||||
Db       1621 ATTCAAGAAAGACCTG 1636
```

RESULT 4
AAC65396

```
ID   AAC65396 standard; cDNA; 1650 BP.
XX
AC   AAC65396;
XX
DT   13-FEB-2001  (first entry)
XX
DE   Human carbohydrate-modifying enzyme cDNA Incyte ID No: 2912330CB1.
XX
KW   Human; carbohydrate-modifying enzyme; CME; antidiabetic;
KW   immunosuppressive; anti-HIV; antiinflammatory; antianaemic;
KW   antiasthmatic; antiarteriosclerotic; antithyroid; hepatotropic;
KW   nephrotropic; antigout; thyromimetic; neuroprotective; osteopathic;
KW   antiarthritic; antipsoriatic; uropathic; ophthalmological;
KW   dermatological; antiulcer; cytostatic; virucide; antibacterial;
KW   fungicide; protozoacide; tranquiliser; vulnerary; diabetes;
KW   autoimmune disorder; inflammatory disorder; infection; ss.
XX
OS   Homo sapiens.
XX
PN   WO200063351-A2.
XX
PD   26-OCT-2000.
XX
PF   20-APR-2000; 2000WO-US10882.
XX
PR   21-APR-1999;   99US-0130383.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Lal P,  Yue H,  Tang YT,  Hillman JL,  Baughn MR,  Yang J;
XX
DR   WPI; 2000-672729/65.
DR   P-PSDB; AAB28677.
XX
PT   Novel carbohydrate modifying enzyme polypeptides and polynucleotides
PT   for diagnosis, treatment, and prevention of carbohydrate metabolism
PT   disorders, autoimmune/inflammatory disorders, and cancer -
XX
PS   Claim 4; Page 75; 75pp; English.
XX
CC   The present cDNA sequence encodes a human carbohydrate-modifying enzyme
CC   (CME). CME polynucleotides and polypeptides are useful for treating and
CC   diagnosing diseases associated with CME such as diabetes,
CC   autoimmune/inflammatory disorders such as AIDS, Addison's disease,
CC   adult respiratory distress syndrome, allergies, anaemia, asthma,
CC   atherosclerosis, autoimmune thyroiditis, bronchitis, cholocystitis,
CC   contact dermatitis, Crohn's disease, emphysema, erythroblastosis fetalis,
CC   glomerulonephritis, Good pasture's syndrome, gout, Grave's disease,
CC   Hashimoto's thyroiditis, multiple sclerosis, myasthenia gravis,
CC   osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis,
CC   Reiter's syndrome, arthritis, scleroderma, sjogren's syndrome, systemic
CC   lupus erythematosus, ulcerative colitis, uveitis, Werner syndrome,
CC   complications of cancer, haemodialysis, and extracorporeal circulation,
CC   viral, bacterial, fungal parasitic, protozoal, and helminthic infections,
CC   trauma, or cancer. CME, or its catalytic or immunogenic fragment, is
CC   useful for drug screening.
XX
SQ   Sequence 1650 BP; 489 A; 330 C; 354 G; 477 T; 0 other;

  Query Match           59.1%;  Score 967.4;  DB 21;  Length 1650;
  Best Local Similarity  75.7%;  Pred. No. 8.2e-267;
  Matches 1228;  Conservative    0;  Mismatches  386;  Indels    9;  Gaps    2;

Qy         23 CATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGCTCCTGCAGCTCTTCTGT---GT 79
              ||||| || ||||   || |   |||  |   |||| ||  | |||||   ||||   | |
Db         28 CATCAGGATGTCTATGAAATGGACTTCAGCTCTTCTCCTGATACAGCTGAGCTGTTACTT 87

Qy         80 TGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTGACATGAGCCATTGGCTTAA 139
              | ||| |||    ||||| || || ||||||||||||   ||  | ||||| ||| | ||
Db         88 TAGCTCTGGGAGTTGTGGAAAGGTGCTGGTGTGGCCCACAGAATTCAGCCACTGGATGAA 147

Qy        140 TGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATGAGGTAACAGTATTGACTCA 199
              | | |||   || || || || ||  |   |||||| |||||||| || |||||| |

Db        148 TATAAAGACAATCCTGGATGAACTTGTCCAGAGAGGTCATGAGGTGACTGTATTGGCATC 207

Qy        200 CTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCATTGAAATTTGAGGTGGTCCA 259
              |||   |   |   | ||   ||  |  || ||| |  | ||||||| ||   |
Db        208 TTCAGCTTCCATTTCTTTCGATCCCAACAGCCCATCTACTCTTAAATTTGAAGTTTATCC 267

Qy        260 TATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTGACCTAGCTCTGA------A 313
              | |  |       | ||| ||     | || ||  |     | || | |  ||
Db        268 TGTATCTTTAACTAAAACTGAGTTTGAGGATATTATCAAGCAGCTGGTTAAGAGATGGGC 327

Qy        314 TGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATGATTTTTTGTTGA 373
              |   | |||       ||  |||  ||   ||  | || || | ||  |  |
Db        328 AGAACTTCCAAAAGACACATTTTGGTCATATTTTTCACAAGTACAAGAAATCATGTGGAC 387

Qy        374 AATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATCAGACGCTTATGAA 433
              | | |  |  |   | | || | | ||| ||    |  | |  ||| ||| |||||||
Db        388 ATTTAATGACATACTTAGAAAGTTCTGTAAGGATATAGTTTCAAATAAGAAACTTATGAA 447

Qy        434 GAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGATTCCCTGTGGAGA 493
              ||| ||||||||  | |  | ||||  | |||  ||| |||| | |||| ||| ||
Db        448 GAAACTACAGGAGTCAAGATTTGATGTTGTTCTTGCAGATGCTGTTTTCCCCTTTGGTGA 507

Qy        494 CCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAATTTCTGTAGGAGG 553
              ||| |||| ||||| |||  | | || |||||   ||  ||  |  | |||   ||
Db        508 GCTGCTGGCCGAGTTACTTAAAATACCCTTTGTCTACAGCCTCCGCTTCTCTCCTGGCTA 567

Qy        554 CAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATGTACCTGTGCCTAT 613
              |   || ||         ||||   |||||      || | |||||||| |||||   |||
Db        568 CGCAATTGAAAAGCATAGTGGAGGACTTCTGTTCCCTCCTTCCTATGTGCCTGTTGTTAT 627

Qy        614 GACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATTCAATGCTTTCAGT 673
              | ||| |||||  |||  |||||| ||  | || ||  |||||||||   ||   |
Db        628 GTCAGAACTAAGTGACCAAATGACTTTCATAGAGAGGGTAAAAAATATGATCTATGTGCT 687

Qy        674 TTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAGAGTTTTATAGTAA 733
              ||  ||  | || ||| | ||    |  |||    |   |||||  |||| || ||| |
Db        688 TTATTTTGAATTTTGGTTCCAAATATTTGACATGAAGAAGTGGGATCAGTTCTACAGTGA 747

Qy        734 GGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTGAGATATGGCTAAT 793
              |   ||||||| |||||||| |||| ||||||  ||| |||||||| ||||||| ||
Db        748 AGTTCTAGGAAGACCCACTACGTTATCTGAGACAATGGCAAAAGCTGACATATGGCTTAT 807

Qy        794 ACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACTTTGAGTTTGTTGG 853
              ||||  || |||||||||| |||||||||| |||  |  ||| ||  ||||||| |||||
Db        808 TCGAAACTACTGGGATTTTCAATTTCCTCACCCACTCTTACCAAATGTTGAGTTCGTTGG 867

Qy        854 AGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAAATTTTGTCCAGAG 913
              |||| | ||||| |||||||||||| |  | || ||||||||||||| | ||||||||||
Db        868 AGGACTCCACTGCAAACCTGCCAAACCCCTACCGAAGGAAATGGAAGAGTTTGTCCAGAG 927

Qy        914 TTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTCAAAATGTTACAGA 973
              || || ||| |||| |||||||||||||||||||||  || |   ||    ||||
Db        928 CTCTGGAGAAAATGGTGTTGTGGTGTTTTCTCTGGGGTCGATGGTCAGTAACACGTCAGA 987

Qy        974 AGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGAAGGTGTTATGGAG 1033
              ||||| ||| ||| | ||||| |||||||||||| ||||||||| |||||  | |||||
Db        988 AGAAAGGGCCAATGTAATTGCATCAGCCCTTGCCAAGATCCCACAAAAGGTTCTGTGGAG 1047

Qy       1034 GTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGTATGATTGGATACC 1093
              |   | || || ||||||   || ||||||  |||||||||||||||  | ||||||||
Db       1048 ATTTGATGGGAATAAACCAGATACTTTAGGACTCAATACTCGGCTGTACAAGTGGATACC 1107

Qy       1094 CCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTCATGGTGGAATGAA 1153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1108 CCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTCATGGTGGAATGAA 1167

Qy       1154 TGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTCCCATATTTGGTGA 1213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1168 TGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTCCCATATTTGGTGA 1227

Qy       1214 TCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAGAAATAAACTTCAA 1273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1228 TCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAGAAATAAACTTCAA 1287
```

```
Qy       1274 AACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTACCGATTCCTCTTA 1333
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1288 AACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTACCGATTCCTCTTA 1347

Qy       1334 TAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTGTAAAGCCCCTAGA 1393
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1348 TAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTGTAAAGCCCCTAGA 1407

Qy       1394 TCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCAAGCACCTGCGATC 1453
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1408 TCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCAAGCACCTGCGATC 1467

Qy       1454 AGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGATTGGGTTCCTGCT 1513
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1468 AGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGATTGGGTTCCTGCT 1527

Qy       1514 GACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTATTTTCCTGTCAAAA 1573
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1528 GACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTATTTTCCTGTCAAAA 1587

Qy       1574 ATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAAATTCAAGAAAGAC 1633
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1588 ATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAAATTCAAGAAAGAC 1647

Qy       1634 CTG 1636
              |||
Db       1648 CTG 1650
```

RESULT 5
AAS91380
ID AAS91380 standard; cDNA; 1766 BP.
XX
AC AAS91380;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #27184.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG27193.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 27184; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1766 BP; 542 A; 359 C; 363 G; 502 T; 0 other;

Query Match 50.2%; Score 821.2; DB 23; Length 1766;
Best Local Similarity 71.9%; Pred. No. 7.4e-225;
Matches 1107; Conservative 0; Mismatches 418; Indels 15; Gaps 2;

```
Qy         94 GTGGGAAAGTCCTGGTGTGGCCCTGTGACATGAGCCATTGGCTTAATGTCAAGGTCATTC 153
              ||||||| ||  || | |||||    ||    || |||||||| ||||| ||| | |||
Db        122 GTGGGAATGTTTTGATTTGGCCAATGGAAGGTAGTCATTGGCTAAATGTTAAGATAATTA 181

Qy        154 TAGAAGAGCTCATAGTGAGAGGCCATGAGGTAACAGTATTGACTCACTCAAAGCCTTCGT 213
              |||| ||||||||     |  |  ||| | || || ||  |   |  |||     |   |
Db        182 TAGATGAGCTCATTAAAAAGGAGCATAATGTGACTGTCCTAGTTGCCTCTGGTGCACTTT 241

Qy        214 TAATTGACTACAGGAAGCCTTCTGCATTGAAATTTGAGGTGGTCCATATGCCACAGGACA 273
              | ||       |          |  | ||| ||||||  |     |  ||||    | ||
Db        242 TCATCACACCAACCTCTAACCCATCTCTGACATTTGAAATATATAAGGTGCCCTTTGGCA 301

Qy        274 GAACAGAAGAAAATGAAATATTTGTTGACCTAGCTCTGAATGTCTTGCCAGGCTTATCAA 333
              |  |  |  | | | | || ||   ||| | | |  ||       ||  |    | ||
Db        302 AAGAAAGAATAGAAGGAGTAATTAAGGACTTCGTTTCGACATGGCTGGAAAATAGACCAT 361

Qy        334 CCTGGCAATCAGTTATAAAATTAAATGATTTTTTTGTTGAAATAAGAGGAACTTTAAAAA 393
              |      | |  ||   | ||| || | |    | |   || |||          || | |
Db        362 CTCCTTCAACCATTTGGAGATTCTATCAGGAGATGGCCAAAGTAATCAAGGACTTCCACA 421

Qy        394 TGATG-----------TGTGAGAGCTTTATCTACAATCAGACGCTTATGAAGAAGCTAC 441
              || ||           |||||  || || |  | || ||   ||| |||   ||||||
Db        422 TGGTGTCTCAGGAGATCTGTGATGGCGTTCTTAAAAACCAACAGCTGATGGCAAAGCTAA 481

Qy        442 AGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGATTCCCTGTGGAGACCTGATGG 501
              || ||| ||| |  || ||  || |    || || ||  |||| ||||| ||  |  | |
Db        482 AGAAAAGCAAGTTTGAAGTCCTGGTGTCTGATCCAGTATTTCCTTGTGGCGATATAGTAG 541

Qy        502 CTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAATTTCTGTAGGAGGCAATATGG 561
              ||      |||| | | || ||| ||  | |  | ||  |||||  ||     |  |||
Db        542 CTTTAAAACTTGGAATTCCATTTATGTACTCCTTGAGGTTTTCTCCAGCCTCAACAGTGG 601

Qy        562 AGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATGTACCTGTGCCTATGACAGGAC 621
              |      |||||||||  | |||   || | |||||||||| ||||    | |  ||| ||
Db        602 AAAAGCACTGTGGGAAGGTACCATACCCTCCTTCCTATGTTCCTGCTGTTTTATCAGAAC 661

Qy        622 TAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATTCAATGCTTTCAGTTTTGTTCC 681
              | || |||  |||| | ||    || ||| ||| |||||   |  | ||        | |
Db        662 TCACCGACCAAATGTCTTTCACTGACAGAATAAGAAATT---TCATCTCCTACCACCTAC 718

Qy        682 ACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAGAGTTTTATAGTAAGGCATTAG 741
              |  || ||   ||       |  | | |||||    |  |||||||| || ||||
Db        719 AGGACTACATGTTTGAAACTCTTTGGAAATCATGGGATTCATACTATAGTAAAGCTTTAG 778

Qy        742 GAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTGAGATATGGCTAATACGAACAT 801
              |||| |||||||| |||||||||||| |||| |||||||| || ||| |||| |||||||
Db        779 GAAGACCCACTACGTTATGTGAGACTATGGGGAAAGCTGAAATTTGGTTAATCCGAACAT 838

Qy        802 ATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACTTTGAGTTTGTTGGAGGATTGC 861
              |||||||||||||||||||  ||||||  ||||||  ||||||||||||||||||||||
```

```
Db        839 ATTGGGATTTTGAATTTCCTCGTCCATACTTACCTAATTTTGAGTTTGTTGGAGGATTGC 898

Qy        862 ACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAAATTTTGTCCAGAGTTCAGGGG 921
              |||| |||||||||||| |||| |||||||||||||| | ||| ||||||| ||||| .
Db        899 ACTGCAAACCTGCCAAACCTTTACCTAAGGAAATGGAAGAATTTATCCAGAGCTCAGGTA 958

Qy        922 AAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTCAAAATGTTACAGAAGAAAAGG 981
              || ||||| |||||||||||||||||| ||| || |  |||| ||||||||||||||
Db        959 AAAATGGTGTTGTGGTGTTTTCTCTGGGATCAATGGTCAAAAACCTTACAGAAGAAAAGG 1018

Qy        982 CTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGAAGGTGTTATGGAGGTACAAAG 1041
              | ||| | ||||| |||||||||||||||| |||||||||| |||||||| |||||||
Db       1019 CCAATCTTATTGCCTCAGCCCTTGCCCAGATTCCACAGAAGGTTTTATGGAGATACAAAG 1078

Qy       1042 GAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGTATGATTGGATACCCCAGAATG 1101
              |||| |||||| |||||||||||  |||||||| ||| | ||||||||||||||||||||
Db       1079 GAAAGAAACCAGCCACATTAGGAAACAATACTCAGCTCTTTGATTGGATACCCCAGAATG 1138

Qy       1102 ATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTCATGGTGGAATGAATGGGATCT 1161
              |||||||||| ||||||||||||||||||||||||||||||||||||  |||||||||
Db       1139 ATCTTCTTGGACATCCCAAAACCAAAGCTTTTATCACTCATGGTGGAACTAATGGGATCT 1198

Qy       1162 ATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTCCCATATTTGGTGATCAGCTTG 1221
              | |||||||||||||| || ||||||||||||||||||||| |||| ||||||| ||
Db       1199 ACGAAGCTATTTACCACGGAGTCCCTATGGTGGGAGTTCCCATGTTTGCTGATCAGCCTG 1258

Qy       1222 ATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAGAAATAAACTTCAAAACTATGA 1281
              ||||||| |||||||||||||||||||||||||||| ||| | ||| | || || ||||
Db       1259 ATAACATTGCTCACATGAAGGCCAAAGGAGCAGCTGTGGAAGTGAACCTAAACACAATGA 1318

Qy       1282 CAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTACCGATTCCTCTTATAAAGAGA 1341
              |||| |  ||||| ||.|| |||||||||||||||||  ||  | ||||||||||||
Db       1319 CAAGTGTGGATTTGCTTAGCGCTTTGAGAACAGTCATTAATGAACCTTCTTATAAAGAGA 1378

Qy       1342 ATGCTATGAGATTATCAAGAATTCACCATGATCAACCTGTAAAGCCCCTAGATCGAGCAG 1401
              |||||||||| |||||||||||||||||||||||||||||||||||||| ||||||||||
Db       1379 ATGCTATGAGGTTATCAAGAATTCACCATGATCAACCTGTAAAGCCCCTGGATCGAGCAG 1438

Qy       1402 TCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCAAGCACCTGCGATCAGCTGCCC 1461
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||    || ||||
Db       1439 TCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCAAGCACCTTCGGGTTGCAGCCC 1498

Qy       1462 ATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGATTGGGTTCCTGCTGACCTGTG 1521
              |||||||||||||||||| || ||||| | ||||| ||||||||| |||||  |||||
Db       1499 ATGACCTCACCTGGTTCCAGTACCACTCTTTGGATGTAATTGGGTTCTTGCTGGTCTGTG 1558

Qy       1522 TGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTATTTTCCTGTCAAAAATTTAATA 1581
              || |||| |||||||| ||| ||| | |||||| ||| ||||||||||||||||  ||
Db       1559 TGACAACGGCTATATTTTTGGTCATACAATGTTGTTTGTTTTCCTGTCAAAAATTTGGTA 1618

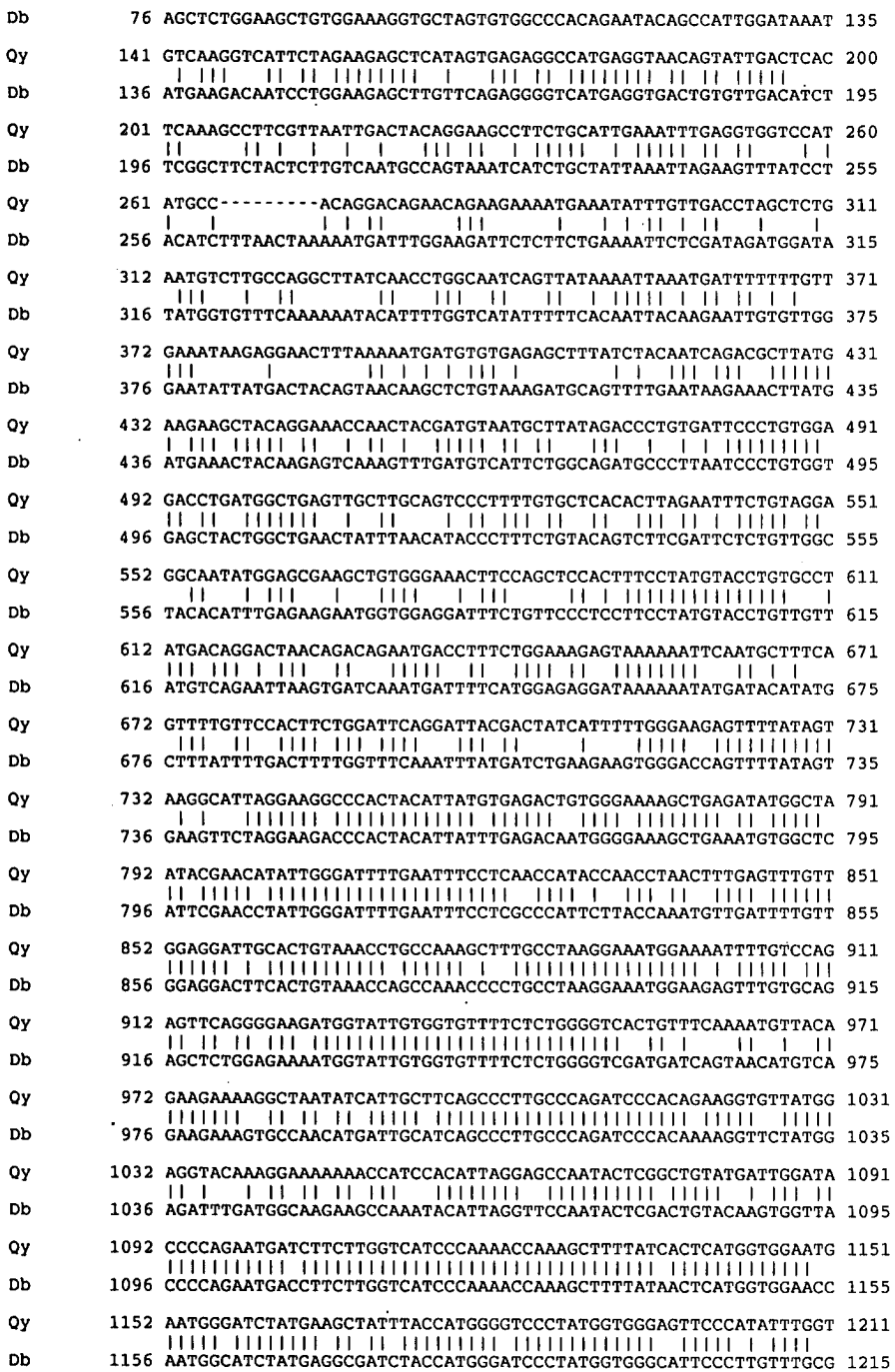
Qy       1582 AAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAAA 1621
              | |   ||||||   |||| || |||||| ||     |||
Db       1619 AGATAGGAAAGAAGAAAAAAAGAGAATAGGTCAAGAAAAA 1658


RESULT 6
ABK84210
ID   ABK84210 standard; cDNA; 2090 BP.
XX
AC   ABK84210;
XX
DT   14-AUG-2002  (first entry)
XX
DE   Human cDNA differentially expressed in granulocytic cells #781.
XX
KW   Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW   viral infection; parasitic infection; protozoal infection;
KW   fungal infection; sterile inflammatory disease; psoriasis;
KW   rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW   cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW   adult respiratory distress syndrome; inflammatory bowel disease;
KW   Crohn's disease; ulcerative colitis; periodontal disease;
KW   granulocyte activation; chronic inflammation; allergy.
XX
OS   Homo sapiens.
XX
PN   WO200228999-A2.
XX
PD   11-APR-2002.
XX
PF   03-OCT-2001; 2001WO-US30821.
XX
PR   03-OCT-2000; 2000US-237189P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Beazer-Barclay Y,  Weissman SM,  Yamaga S,  Vockley J;
XX
DR   WPI; 2002-435328/46.
XX
PT   Detecting granulocyte activation by detecting differential expression
PT   of genes associated with granulocyte activation, which serves as
PT   diagnostic markers that is useful for monitoring disease states and
PT   drug toxicity -
XX
PS   Claim 1; SEQ ID No 781; 114pp; English.
XX
CC   The invention relates to detecting (M1) granulocyte (GC) activation
CC   (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC   DNA chip analysis as given in the specification, and comparing
CC   the expression level to an expression level in an unactivated
CC   GC, where differential expression of Gs is indicative of GCA.
CC   Also included are modulating (M2) GA by contacting GC with an agent
CC   that alters the expression of at least one gene in Gs; (2) screening (M3)
CC   for an agent capable of modulating GCA or an inflammation (especially
CC   chronic) in a tissue, an allergic response in a subject, exposure of a
CC   subject to a pathogen or sterile inflammatory disease using the
CC   gene expression profile; (3) detecting (M4) an inflammation (especially
CC   chronic) in a tissue, an allergic response in a subject, exposure of a
CC   subject to a pathogen or sterile inflammatory disease, by detecting the
CC   level of expression in a sample of the tissue of gene(s) from Gs, where
CC   the level of expression of the gene is indicative of inflammation;
CC   (4) treating (M5) an inflammation (especially chronic) or in a tissue,
CC   an allergic response in a subject, exposure of a subject to a pathogen
CC   or sterile inflammatory disease, by contacting a tissue having
CC   inflammation with an agent that modulates the expression of gene(s)
CC   from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
CC   modulating GA; M3 is useful for screening an agent capable of modulating
CC   GCA preferably in an inflammation in a tissue; M4 is useful for
CC   detecting an inflammation (especially chronic) in a tissue, an allergic
CC   response in a subject, exposure of a subject to a pathogen or sterile
CC   inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
CC   glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
CC   reperfusion injury, ARDS, adult respiratory distress syndrome,
CC   inflammatory bowel disease, Crohn's disease, ulcerative colitis,
CC   periodontal disease; also bacterial infection, viral infection,
CC   parasitic infection, protozoal infection, fungal infection and M5 is
CC   useful for treating one of the above conditions. The present
CC   sequence represents a gene differentially expressed in granulocytes.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2090 BP; 649 A; 384 C; 435 G; 622 T; 0 other;

  Query Match             47.8%;  Score 782;  DB 24;  Length 2090;
  Best Local Similarity   68.8%;  Pred. No. 1.4e-213;
  Matches 1108;  Conservative    0;  Mismatches  490;  Indels   12;  Gaps    2;

Qy         24 ATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGCTCCTGCAGCTCTTCTGT---GTT 80
              | || || ||||   || |   | |  || ||||||||  | ||||||   |||    ||
Db         16 ACCAGGATGTCTCTGAAATGGACGTCAGTCTTTCTGCTGATACAGCTCAGTTGTTACTTT 75

Qy         81 GGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTGACATGAGCCATTGGCTTAAT 140
                ||| ||||  |||||| || || || |||||||||   ||    ||||||||| | |||
```





RESULT 7
ABN96795
ID ABN96795 standard; DNA; 2090 BP.
XX
AC ABN96795;
XX
DT 13-AUG-2002 (first entry)
XX
DE Gene #3293 used to diagnose liver cancer.
XX
KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW metastatic liver tumour; cytostatic; expression profile; disease state;
KW disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
OS Homo sapiens.
XX
PN WO200229103-A2.
XX
PD 11-APR-2002.
XX
PF 02-OCT-2001; 2001WO-US30589.
XX
PR 02-OCT-2000; 2000US-237054P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX
DR WPI; 2002-426119/45.
XX
PT Diagnosing and detecting the progression of liver cancer,
PT hepatocellular carcinoma or metastatic liver tumor in a patient,
PT involves detecting the level of expression of two or more genes in a
PT liver tissue sample -
XX
PS Claim 1; SEQ ID NO 3293; 298pp; English.
XX
CC The invention relates to a novel method for diagnosing and detecting the
CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC tumour in a patient, and differentiating metastatic liver cancer from
CC hepatocellular carcinoma in a patient, involving detecting the level of
CC expression of two or more genes represented in ABN93503-ABN97455 in a
CC tissue sample. The method of the invention has hepatotropic, and

```
CC   cytostatic activity. The method is useful for diagnosing and detecting
CC   the progression of liver cancer, hepatocellular carcinoma and metastatic
CC   liver carcinoma in a patient. The method is useful for identifying
CC   expression profiles which serve as useful diagnostic markers as well as
CC   markers that can be used to monitor disease states, disease progression,
CC   drug toxicity, drug efficacy and drug metabolism.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2090 BP; 649 A; 384 C; 435 G; 622 T; 0 other;

  Query Match            47.8%;  Score 782;  DB 24;  Length 2090;
  Best Local Similarity  68.8%;  Pred. No. 1.4e-213;
  Matches 1108;  Conservative    0;  Mismatches  490;  Indels   12;  Gaps    2;

Qy         24 ATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGCTCCTGCAGCTCTTCTGT---GTT 80
Db         16 ACCAGGATGTCTCTGAAATGGACGTCAGTCTTTCTGCTGATACAGCTCAGTTGTTACTTT 75

Qy         81 GGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTGACATGAGCCATTGGCTTAAT 140
Db         76 AGCTCTGGAAGCTGTGGAAAGGTGCTAGTGTGGCCCACAGAATACAGCCATTGGATAAAT 135

Qy        141 GTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATGAGGTAACAGTATTGACTCAC 200
Db        136 ATGAAGACAATCCTGGAAGAGCTTGTTCAGAGGGGTCATGAGGTGACTGTGTTGACATCT 195

Qy        201 TCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCATTGAAATTTGAGGTGGTCCAT 260
Db        196 TCGGCTTCTACTCTTGTCAATGCCAGTAAATCATCTGCTATTAAATTAGAAGTTTATCCT 255

Qy        261 ATGCC---------ACAGGACAGAACAGAAGAAAATGAAATATTTGTTGACCTAGCTCTG 311
Db        256 ACATCTTTAACTAAAAATGATTTGGAAGATTCTCTTCTGAAAATTCTCGATAGATGGATA 315

Qy        312 AATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATGATTTTTTTGTT 371
Db        316 TATGGTGTTTCAAAAAATACATTTTGGTCATATTTTTCACAATTACAAGAATTGTGTTGG 375

Qy        372 GAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATCAGACGCTTATG 431
Db        376 GAATATTATGACTACAGTAACAAGCTCTGTAAAGATGCAGTTTTGAATAAGAAACTTATG 435

Qy        432 AAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGATTCCCTGTGGA 491
Db        436 ATGAAACTACAAGAGTCAAAGTTTGATGTCATTCTGGCAGATGCCCTTAATCCCTGTGGT 495

Qy        492 GACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAATTTCTGTAGGA 551
Db        496 GAGCTACTGGCTGAACTATTTAACATACCCTTTCTGTACAGTCTTCGATTCTCTGTTGGC 555

Qy        552 GGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATGTACCTGTGCCT 611
Db        556 TACACATTTGAGAAGAATGGTGGAGGATTTCTGTTCCCTCCTTCCTATGTACCTGTTGTT 615

Qy        612 ATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATTCAATGCTTTCA 671
Db        616 ATGTCAGAATTAAGTGATCAAATGATTTTCATGGAGAGGATAAAAAATATGATACATATG 675

Qy        672 GTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAGAGTTTTATAGT 731
Db        676 CTTTATTTTGACTTTTGGTTTCAAATTTATGATCTGAAGAAGTGGGACCAGTTTTATAGT 735

Qy        732 AAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTGAGATATGGCTA 791
Db        736 GAAGTTCTAGGAAGACCCACTACATTATTTGAGACAATGGGGAAAGCTGAAATGTGGCTC 795

Qy        792 ATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACTTTGAGTTTGTT 851
Db        796 ATTCGAACCTATTGGGATTTTGAATTTCCTCGCCCATTCTTACCAAATGTTGATTTTGTT 855

Qy        852 GGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAAATTTTGTCCAG 911
Db        856 GGAGGACTTCACTGTAAACCAGCCAAACCCCTGCCTAAGGAAATGGAAGAGTTTGTGCAG 915

Qy        912 AGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTCAAAATGTTACA 971
Db        916 AGCTCTGGAGAAAATGGTATTGTGGTGTTTTCTCTGGGGTCGATGATCAGTAACATGTCA 975

Qy        972 GAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGAAGGTGTTATGG 1031
Db        976 GAAGAAAGTGCCAACATGATTGCATCAGCCCTTGCCCAGATCCCACAAAAGGTTCTATGG 1035

Qy       1032 AGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGTATGATTGGATA 1091
Db       1036 AGATTTGATGGCAAGAAGCCAAATACATTAGGTTCCAATACTCGACTGTACAAGTGGTTA 1095

Qy       1092 CCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTCATGGTGGAATG 1151
Db       1096 CCCCAGAATGACCTTCTTGGTCATCCCAAAACCAAAGCTTTTATAACTCATGGTGGAACC 1155

Qy       1152 AATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTCCCATATTTGGT 1211
Db       1156 AATGGCATCTATGAGGCGATCTACCATGGGATCCCTATGGTGGGCATTCCCTTGTTTGCG 1215

Qy       1212 GATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAGAAATAAACTTC 1271
Db       1216 GATCAACATGATAACATTGCTCACATGAAAGCCAAGGGAGCAGCCCTCAGTGTGGACATC 1275

Qy       1272 AAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTACCGATTCCTCT 1331
Db       1276 AGGACCATGTCAAGTAGAGATTTGCTCAATGCATTGAAGTCAGTCATTAATGACCCTGTC 1335

Qy       1332 TATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTGTAAAGCCCCTA 1391
Db       1336 TATAAAGAGAATGTCATGAAATTATCAAGAATTCATCATGACCAACCAATGAAGCCCCTG 1395

Qy       1392 GATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCAAGCACCTGCGA 1451
Db       1396 GATCGAGCAGTCTTCTGGATTGAGTTTGTCATGCGCCACAAAGGAGCCAAGCACCTTCGA 1455

Qy       1452 TCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGATTGGGTTCCTG 1511
Db       1456 GTCGCAGCTCACAACCTCACCTGGATCCAGTACCACTCTTTGGATGTGATAGCATTCCTG 1515

Qy       1512 CTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTATTTTCCTGTCAA 1571
Db       1516 CTGGCCTGCGTGGCAACTGTGATATTTATCATCACAAAATTTTGCCTGTTTTGTTTCCGA 1575

Qy       1572 AAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAAA 1621
Db       1576 AAGCTTGCCAAAACAGGAAAGAAGAAGAAAAGAGATTAGTTATATCAAAA 1625

RESULT 8
AAZ95206
ID   AAZ95206 standard; DNA; 1976 BP.
XX
AC   AAZ95206;
XX
DT   05-JUN-2000  (first entry)
XX
DE   Human UDP-glucuronosyltransferase 2B15 nucleotide sequence.
XX
KW   UDP-glucuronosyltransferase 2B15; UGT2B15; polymorphism; metabolism;
KW   drug interaction; detect; human; single nucleotide polymorphism;
KW   SNPs; ds.
XX
OS   Homo sapiens.
XX
PN   WO200006776-A1.
XX
PD   10-FEB-2000.
XX
PF   22-JUL-1999;   99WO-US16675.
```

XX
PR 28-JUL-1998; 98US-0094391.
XX
PA (AXYS-) AXYS PHARM INC.
XX
PI Galvin M, Miller A, Penny L, Riedy M;
XX
DR WPI; 2000-195321/17.
DR P-PSDB; AAY78935.
XX
PT Novel human UDP-glucuronosyltransferase sequence, polymorphisms for
PT genotyping individuals to predict rate of metabolism of substrates and
PT for identifying potential drug interactions -
XX
PS Disclosure; Page 56-59; 72pp; English.
XX
CC This sequence represents the human UDP-glucuronosyltransferase 2B15
CC (UGT2B15) gene. UDP-glucuronosyltransferase (UGTs) are a family of
CC enzymes that catalyse the glucuronic acid conjugation of a wide range of
CC endogenous and exogenous substrates. The UGT2B gene subfamily encode
CC steroid metabolizing isoforms in the liver. Alteration of the expression
CC or function of UGTs may effect drug metabolism. The invention relates to
CC non-chromosomal nucleic acid molecules, which comprise human UGT2B
CC sequence polymorphisms (see AAZ95051-Z95110). Probes which detect the
CC UGT2B locus polymorphisms can be used to detect altered UGT2B metabolism
CC of a substrate in an individual. The nucleic acid molecules comprising a
CC human UGT2B sequence polymorphism can be used in screening assays for
CC genotyping individuals, also to predict their rate of metabolism of
CC UGT2B substrate, potential drug-drug interactions and adverse side
CC effects. The polymorphisms can be used as single nucleotide polymorphisms
CC (SNPs) for detecting genetic linkage related to phenotypic variation in
CC activity or expression of UGT2B protein. The polymorphism containing
CC nucleic acid molecules may also be used for generating genetically
CC modified non-human animals and for obtaining site specific gene
CC modification in cell lines.
XX
SQ Sequence 1976 BP; 594 A; 368 C; 419 G; 595 T; 0 other;

Query Match 47.7%; Score 780.4; DB 21; Length 1976;
Best Local Similarity 68.8%; Pred. No. 4e-213;
Matches 1107; Conservative 0; Mismatches 491; Indels 12; Gaps 2;

```
Qy    24 ATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGCTCCTGCAGCTCTTCTGT---GTT  80
Db     5 ACCAGGATGTCTCTGAAATGGACGTCAGTCTTTCTGCTGATACAGCTCAGTTGTTACTTT  64

Qy    81 GGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTGACATGAGCCATTGGCTTAAT 140
Db    65 AGCTCTGGAAGCTGTGGAAAGGTGCTAGTGTGGCCCACAGAATACAGCCATTGGATAAAT 124

Qy   141 GTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATGAGGTAACAGTATTGACTCAC 200
Db   125 ATGAAGACAATCCTGGAAGAGCTTGTTCAGAGGGGTCATGAGGTGACTGTGTTGACATCT 184

Qy   201 TCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCATTGAAATTTGAGGTGGTCCAT 260
Db   185 TCGGCTTCTACTCTTGTCAATGCCAGTAAATCATCTGCTATTAAATTAGAAGTTTATCCT 244

Qy   261 ATGCC---------ACAGGACAGAACAGAAGAAAATGAAATATTTGTTGACCTAGCTCTG 311
Db   245 ACATCTTTAACTAAAAATGATTTGGAAGATTCTCTTCTGAAAATTCTCGATAGATGGATA 304

Qy   312 AATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATGATTTTTTTGTT 371
Db   305 TATGGTGTTTCAAAAAATACATTTTGGTCATATTTTTCACAATTACAAGAATTGTGTTGG 364

Qy   372 GAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATCAGACGCTTATG 431
Db   365 GAATATTATGACTACAGTAACAAGCTCTGTAAAGATGCAGTTTTGAATAAGAAACTTATG 424

Qy   432 AAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGATTCCCTGTGGA 491
Db   425 ATGAAACTACAAGAGTCAAAGTTTGATGTCATTCTGGCAGATGCCCTTAATCCCTGTGGT 484

Qy   492 GACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAATTTCTGTAGGA 551
Db   485 GAGCTACTGGCTGAACTATTTAACATACCCTTTCTGTACAGTCTTCGATTCTCTGTTGGC 544

Qy   552 GGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATGTACCTGTGCCT 611
Db   545 TACACATTTGAGAAGAATGGTGGAGGATTTCTGTTCCCTCCTTCCTATGTACCTGTTGTT 604

Qy   612 ATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATTCAATGCTTTCA 671
Db   605 ATGTCAGAATTAAGTGATCAAATGATTTTCATGGAGAGGATAAAAAATATGATACATATG 664

Qy   672 GTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAGAGTTTTATAGT 731
Db   665 CTTTATTTTGACTTTTGGTTTCAAATTTATGATCTGAAGAAGTGGGACCAGTTTTATAGT 724

Qy   732 AAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTGAGATATGGCTA 791
Db   725 GAAGTTCTAGGAAGACCCACTACATTATTTGAGACAATGGGGAAAGCTGAAATGTGGCTC 784

Qy   792 ATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACTTTGAGTTTGTT 851
Db   785 ATTCGAACCTATTGGGATTTTGAATTTCCTCGCCCATTCTTACCAAATGTTGATTTTGTT 844

Qy   852 GGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAAATTTTGTCCAG 911
Db   845 GGAGGACTTCACTGTAAACCAGCCAAACCCCTGCCTAAGGAAATGGAAGAGTTTGTGCAG 904

Qy   912 AGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTCAAAATGTTACA 971
Db   905 AGCTCTGGAGAAAATGGTATTGTGGTGTTTTCTCTGGGGTCGATGATCAGTAACATGTCA 964

Qy   972 GAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGAAGGTGTTATGG 1031
Db   965 GAAGAAAGTGCCAACATGATTGCATCAGCCCTTGCCCAGATCCCACAAAAGGTTCTATGG 1024

Qy  1032 AGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGTATGATTGGATA 1091
Db  1025 AGATTTGATGGCAAGAAGCCAAATACTTTAGGTTCCAATACTCGACTGTACAAGTGGTTA 1084

Qy  1092 CCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTCATGGTGGAATG 1151
Db  1085 CCCCAGAATGACCTTCTTGGTCATCCCAAAACCAAAGCTTTTATAACTCATGGTGGAACC 1144

Qy  1152 AATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTCCCATATTTGGT 1211
Db  1145 AATGGCATCTATGAGGCGATCTACCATGGGATCCCTATGGTGGGCATTCCCTTGTTTGCG 1204

Qy  1212 GATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAGAAATAAACTTC 1271
Db  1205 GATCAACATGATAACATTGCTCACATGAAAGCCAAGGGAGCAGCCCTCAGTGTGGACATC 1264

Qy  1272 AAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTACCGATTCCTCT 1331
Db  1265 AGGACCATGTCAAGTAGAGATTTGCTCAATGCATTGAAGTCAGTCATTAATGACCCTGTC 1324

Qy  1332 TATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTGTAAAGCCCCTA 1391
Db  1325 TATAAAGAGAATGTCATGAAATTATCAAGAATTCATCATGACCAACCAATGAAGCCCCTG 1384

Qy  1392 GATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCAAGCACCTGCGA 1451
Db  1385 GATCGAGCAGTCTTCTGGATTGAGTTTGTCATGCGCCACAAAGGAGCCAAGCACCTTCGA 1444

Qy  1452 TCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGATTGGGTTCCTG 1511
Db  1445 GTCGCAGCTCACAACCTCACCTGGATCCAGTACCACTCTTTGGATGTGATAGCATTCCTG 1504

Qy  1512 CTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTATTTTCCTGTCAA 1571
Db  1505 CTGGCCTGCGTGGCAACTGTGATATTTATCATCACAAAATTTTGCCTGTTTTGTTTCCGA 1564
```

```
Qy       1572 AAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAAA 1621
Db       1565 AAGCTTGCCAAAACAGGAAAGAAGAAGAAAAGAGATTAGTTATATCAAAA 1614


RESULT 9
AAV15900
ID   AAV15900 standard; cDNA; 2107 BP.
XX
AC   AAV15900;
XX
DT   26-MAY-1998  (first entry)
XX
DE   Uridine diphospho-glucuronosyltransferase 2B17 (UGT2B17) encoding cDNA.
XX
KW   Uridine diphospho-glucuronosyltransferase 2B17; UGT2B17; catalyse;
KW   androsterone; androsterone-glucuronic acid; androgen; enzyme; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           1..51
FT                   /*tag= a
FT   CDS             52..1644
FT                   /*tag= b
FT                   /product= "UGTB17 enzyme"
FT   3'UTR           1645..2107
FT                   /*tag= c
XX
PN   WO9744466-A1.
XX
PD   27-NOV-1997.
XX
PF   16-MAY-1997;   97WO-CA00328.
XX
PR   17-MAY-1996;   96US-0649319.
XX
PA   (ENDO-) ENDORECHERCHE INC.
XX
PI   Beaulieu M,  Belanger A,  Hum DW,  Levesque E;
XX
DR   WPI; 1998-018520/02.
DR   P-PSDB; AAW47126.
XX
PT   DNA encoding uridine di:phospho:glucuronosyl:transferase 2B17 -
PT   which catalyses conversion of androsterone to
PT   androsterone-glucuronic acid
XX
PS   Claim 15; Pages 4-6; 53pp; English.
XX
CC   This cDNA encodes an enzyme uridine di-phosphoglucuronosyltransferase
CC   2B17 (UGT2B17). This novel enzyme catalyses the conversion of
CC   androsterone to androsterone-glucuronic acid. The UGT2B17 can be used to
CC   detect anti-UGT2B17 antibodies. The antibody can be used to detect a
CC   localised concentration of UGT2B17 or an alteration in androgen activity.
CC   The UGT2B17 can also be used to alter the concentration of an androgenic
CC   compound in a tissue, specifically dihydrotestosterone. An isolated
CC   nucleotide sequence comprising at least 30 consecutive nucleotides from
CC   the coding region of the 2107 base pair sequence, or its complement can
CC   be used to block the synthesis of UGT2B17, e.g. an expression disrupting
CC   sense or antisense fragment, or as a probe for a UGT2B17 coding sequence.
XX
SQ   Sequence 2107 BP; 657 A; 382 C; 424 G; 644 T; 0 other;

  Query Match             47.4%;  Score 775.6;  DB 19;  Length 2107;
  Best Local Similarity   68.6%;  Pred. No. 9.7e-212;
  Matches 1104;  Conservative    0;  Mismatches  494;  Indels   12;  Gaps    2;

Qy         24 ATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGCTCCTGCAGCTCTTCTGT---GTT 80
Db         46 ACCAGGATGTCTCTGAAATGGATGTCAGTCTTTCTGCTGATGCAGCTCAGTTGTTACTTT 105

Qy         81 GGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTGACATGAGCCATTGGCTTAAT 140
Db        106 AGCTCTGGGAGTTGTGGAAAGGTGCTGGTGTGGCCCACAGAATACAGCCATTGGATAAAT 165

Qy        141 GTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATGAGGTAACAGTATTGACTCAC 200
Db        166 ATGAAGACAATCCTGGAAGAGCTTGTTCAGAGGGGTCATGAGGTGATTGTGTTGACATCT 225

Qy        201 TCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCATTGAAATTTGAGGTGGTCCAT 260
Db        226 TCGGCTTCTATTCTTGTCAATGCCAGTAAATCATCTGCTATTAAATTAGAAGTTTATCCT 285

Qy        261 ATGCCACAGGACAGAACAGAAGAAAATGAAATATTT---------GTTGACCTAGCTCTG 311
Db        286 ACATCTTTAACTAAAAATGATTTGGAAGATTTTTTTATGAAAATGTTCGATAGATGGACA 345

Qy        312 AATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATGATTTTTTGTT 371
Db        346 TATAGTATTTCAAAAAATACATTTTGGTCATATTTTTCACAACTACAAGAATTGTGTTGG 405

Qy        372 GAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATCAGACGCTTATG 431
Db        406 GAATATTCTGACTATAATATAAAGCTCTGTGAAGATGCAGTTTTGAACAAGAAACTTATG 465

Qy        432 AAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGATTCCCTGTGGA 491
Db        466 AGAAAACTACAAGAGTCAAAATTTGATGTCCTTCTGGCAGATGCCGTTAATCCCTGTGGT 525

Qy        492 GACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAATTTCTGTAGGA 551
Db        526 GAGCTGCTGGCTGAACTACTTAACATACCCTTTCTGTACAGTCTCCGCTTCTCTGTTGGC 585

Qy        552 GGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATGTACCTGTGCCT 611
Db        586 TACACAGTTGAGAAGAATGGTGGAGGATTTCTGTTCCCTCCTTCCTATGTACCTGTTGTT 645

Qy        612 ATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATTCAATGCTTTCA 671
Db        646 ATGTCAGAATTAAGTGATCAAATGATTTTCATGGAGAGGATAAAAAATATGATATATATG 705

Qy        672 GTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAGAGTTTTATAGT 731
Db        706 CTTTATTTTGACTTTTGGTTTCAAGCATATGATCTGAAGAAGTGGGACCAGTTTTATAGT 765

Qy        732 AAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTGAGATATGGCTA 791
Db        766 GAAGTTCTAGGAAGACCCACTACATTATTTGAGACAATGGGGAAAGCTGAAATGTGGCTC 825

Qy        792 ATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACTTTGAGTTTGTT 851
Db        826 ATTCGAACCTATTGGGATTTTGAATTTCCTCGCCCATTCTTACCAAATGTTGATTTTGTT 885

Qy        852 GGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAAATTTTGTCCAG 911
Db        886 GGAGGACTTCACTGTAAACCAGCCAAACCCTTGCCTAAGGAAATGGAAGAGTTTGTGCAG 945

Qy        912 AGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTCAAAATGTTACA 971
Db        946 AGCTCTGGAGAAAATGGTATTGTGGTGTTTTCTCTGGGGTCGATGATCAGTAACATGTCA 1005

Qy        972 GAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGAAGGTGTTATGG 1031
Db       1006 GAAGAAAGTGCCAACATGATTGCATCAGCCCTTGCCCAGATCCCACAAAAGGTTCTATGG 1065

Qy       1032 AGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGTATGATTGGATA 1091
Db       1066 AGATTTGATGGCAAGAAGCCAAATACTTTAGGTTCCAATACTCGACTGTATAAGTGGTTA 1125

Qy       1092 CCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTCATGGTGGAATG 1151
Db       1126 CCCCAGAATGACCTTCTTGGTCATCCCAAAACCAAAGCTTTTATAACTCATGGTGGAACC 1185

Qy       1152 AATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTCCCATATTTGGT 1211
```

```
Db       1186 AATGGCATCTATGAGGCGATCTACCATGGGATCCCTATGGTGGGCATTCCCTTGTTTGCG 1245

Qy       1212 GATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAGAAATAAACTTC 1271

Db       1246 GATCAACATGATAACATTGCTCACATGAAAGCCAAGGGAGCAGCCCTCAGTGTGGACATC 1305

Qy       1272 AAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTACCGATTCCTCT 1331

Db       1306 AGGACCATGTCAAGTAGAGATTTGCTCAATGCATTGAAGTCAGTCATTAATGACCCTATC 1365

Qy       1332 TATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTGTAAAGCCCCTA 1391

Db       1366 TATAAAGAGAATATCATGAAATTATCAAGAATTCATCATGATCAACCGGTGAAGCCCCTG 1425

Qy       1392 GATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCAAGCACCTGCGA 1451

Db       1426 GATCGAGCAGTCTTCTGGATTGAGTTTGTCATGCGCCATAAAGGAGCCAAGCACCTTCGG 1485

Qy       1452 TCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGATTGGGTTCCTG 1511

Db       1486 GTCGCAGCCCACAACCTCACCTGGATCCAGTACCACTCTTTGGATGTGATAGCATTCCTG 1545

Qy       1512 CTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTATTTTCCTGTCAA 1571

Db       1546 CTGGCCTGCGTGGCAACTATGATATTTATGATCACAAAATGTTGCCTGTTTTGTTTCCGA 1605

Qy       1572 AAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAAA 1621

Db       1606 AAGCTTGCCAAAACAGGAAAGAAGAAGAAAAGGGATTAGTTATATCAAAA 1655
```

RESULT 10
ABN96789
ID ABN96789 standard; DNA; 2123 BP.
XX
AC ABN96789;
XX
DT 13-AUG-2002 (first entry)
XX
DE Gene #3287 used to diagnose liver cancer.
XX
KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW metastatic liver tumour; cytostatic; expression profile; disease state;
KW disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
OS Homo sapiens.
XX
PN WO200229103-A2.
XX
PD 11-APR-2002.
XX
PF 02-OCT-2001; 2001WO-US30589.
XX
PR 02-OCT-2000; 2000US-237054P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX
DR WPI; 2002-426119/45.
XX
PT Diagnosing and detecting the progression of liver cancer,
PT hepatocellular carcinoma or metastatic liver tumor in a patient,
PT involves detecting the level of expression of two or more genes in a
PT liver tissue sample
XX
PS Claim 1; SEQ ID NO 3287; 298pp; English.
XX
CC The invention relates to a novel method for diagnosing and detecting the
CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC tumour in a patient, and differentiating metastatic liver cancer from
CC hepatocellular carcinoma in a patient, involving detecting the level of
CC expression of two or more genes represented in ABN93503-ABN97455 in a
CC tissue sample. The method of the invention has hepatotropic, and
CC cytostatic activity. The method is useful for diagnosing and detecting
CC the progression of liver cancer, hepatocellular carcinoma and metastatic
CC liver carcinoma in a patient. The method is useful for identifying
CC expression profiles which serve as useful diagnostic markers as well as
CC markers that can be used to monitor disease states, disease progression,
CC drug toxicity, drug efficacy and drug metabolism.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2123 BP; 676 A; 389 C; 427 G; 631 T; 0 other;

Query Match 46.8%; Score 766.2; DB 24; Length 2123;
Best Local Similarity 68.4%; Pred. No. 4.9e-209;
Matches 1078; Conservative 0; Mismatches 489; Indels 9; Gaps 1;

```
Qy         55 TTCTGCTCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGC 114

Db          8 TGCTGATACAGCTCAGTTGTTACTTTAGCTCTGGAAGCTGTGGAAAGGTGCTAGTGTGGC 67

Qy        115 CCTGTGACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAG 174

Db         68 CCACAGAATACAGCCATTGGATAAATATGAAGACAATCCTGGAAGAGCTTGTTCAGAGGG 127

Qy        175 GCCATGAGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTT 234

Db        128 GTCATGAGGTGACTGTGTTGACATCTTCGGCTTCTACTCTTGTCAATGCCAGTAAATCAT 187

Qy        235 CTGCATTGAAATTTGAGGTGGTCCATATGCC---------ACAGGACAGAACAGAAGAAA 285

Db        188 CTGCTATTAAATTAGAAGTTTATCCTACATCTTTAACTAAAAATTATTTGGAAGATTCTC 247

Qy        286 ATGAAATATTTGTTGACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAG 345

Db        248 TTCTGAAAATTCTCGATAGATGGATATATGGTGTTTCAAAAAATACATTTTGGTCATATT 307

Qy        346 TTATAAAATTAAATGATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGA 405

Db        308 TTTCACAATTACAAGAATTGTGTTGGGCATATTATGACTACAGTAACAAGCTCTGTAAAG 367

Qy        406 GCTTTATCTACAATCAGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGC 465

Db        368 ATGCAGTTTTGAATAAGAAACTTATGATGAAACTACAAGAGTCAAGATTTGATGTCATTC 427

Qy        466 TTATAGACCCTGTGATTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTG 525

Db        428 CTGGAGATCCTGTTTTCCCCTGTGGTGAGCTACTGTCAGAGCTACTTAACATACCCTTTG 487

Qy        526 TGCTCACACTTAGAATTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAG 585

Db        488 TCTACAGGTCTCGCATCTCTCGTGGCTACACAATTGAGAAGAATGGTGGAGGATTTCTGT 547

Qy        586 CTCCACTTTCCTATGTACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGG 645

Db        548 TCCCTCCTTCCTATGTACCTGTTGTTATGTCAGAATTAATTGATCAAATGATTTTCATGG 607

Qy        646 AAAGAGTAAAAAATTCAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACT 705

Db        608 AGAGGATAAAAAATATGATACATATGCTTTATTTTGACTTTTGGTTTCAAATTTATGATC 667

Qy        706 ATCATTTTTGGGAAGAGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGA 765

Db        668 TGAAGAAGTGGGACCAGTTTTATAGTGAAGTTCTAGGAAGACCCACTACATTATTTGAGA 727

Qy        766 CTGTGGGAAAAGCTGAGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAAC 825

Db        728 CAATGGGGAAAGCTGAAATGTGGCTCATTCGAACCTATTGGGATTTTGAATTTCCTCGCC 787

Qy        826 CATACCAACCTAACTTTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGC 885

Db        788 CATTCTTACCAAATGTTGATTTTGTTGGAGGACTTCACTGTAAACCAGCCAAACCCCTGC 847
```

```
Qy        886 CTAAGGAAATGGAAAATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTC 945
              |||||||| ||||| | ||||| ||||| || || ||| ||||||||||||||||||||
Db        848 CTAAGGAATTGGAAGAGTTTGTGCAGAGCTCTGGAGAAAATGGTATTGTGGTGTTTTCTC 907

Qy        946 TGGGGTCACTGTTTCAAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTG 1005
              |||||||  || |    ||  |  |||||||||  || || || ||||| ||||||||||
Db        908 TGGGGTCGATGATCAGTAACATGTCAGAAGAAAGTGCCAACATGATTGCATCAGCCCTTG 967

Qy       1006 CCCAGATCCCACAGAAGGTGTTATGGAGGTACAAAGGAAAAAAAACCATCCACATTAGGAG 1065
              ||||||||||||  |||||  |||||||  |    | || || || |||   || |||||
Db        968 CCCAGATCCCACAAAAGGTTCTATGGAGATTTGATGGCAAGAAGCCAAATACTTTAGGTT 1027

Qy       1066 CCAATACTCGGCTGTATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCA 1125
              |||||||||| |||||  | |||  ||||||||||||| ||||||||||||||||||||
Db       1028 CCAATACTCGACTGTACAAGTGGTTACCCCAGAATGACCTTCTTGGTCATCCCAAAACCA 1087

Qy       1126 AAGCTTTTATCACTCATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCC 1185
              |||||||||| ||||||||||||  ||||| |||||||| || || ||||||||| |||
Db       1088 AAGCTTTTATAACTCATGGTGGAACCAATGGCATCTATGAGGCGATCTACCATGGGATCC 1147

Qy       1186 CTATGGTGGGAGTTCCCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCA 1245
              |||||||||  ||||| | ||||  |||||   |||||||| |||||||||| ||||
Db       1148 CTATGGTGGGCATTCCCTTGTTTGCGGATCAAGATGATAACATTGCTCACATGAAAGCCA 1207

Qy       1246 AAGGAGCAGCTGTAGAAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTT 1305
              | ||||||||  |      |  || |||  || ||| ||||   |||||| || |  || |
Db       1208 AGGGAGCAGCCCTCAGTGTGGACATCAGGACCATGTCAAGTAGAGATTTGCTCAATGCAT 1267

Qy       1306 TGAGAACAGTCATTACCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTC 1365
              |||   |||||||||  ||  | .  |||||||||||||||||| ||||||||||||
Db       1268 TGAAGTCAGTCATTAATGACCCAGTCTATAAAGAGAATGCTATGAAATTATCAAGAATTC 1327

Qy       1366 ACCATGATCAACCTGTAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGC 1425
              | ||||| |||||  | ||||||||  |||||||||||||||||||| ||||||||||||
Db       1328 ATCATGACCAACCAATGAAGCCCCTGGATCGAGCAGTCTTCTGGATTGAGTTTGTCATGC 1387

Qy       1426 GCCACAAAGGAGCCAAGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACT 1485
              |||||||||||||||||||| |||   || || ||  |||||||||| ||||| ||
Db       1388 GCCACAAAGGAGCCAAGCACCTTCGAGTCGCAGCTCACAACCTCACCTGGATCCAGTACC 1447

Qy       1486 ACTCTATAGATGTGATTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCA 1545
              ||||| | ||||||||  |  |||||||||  |||| ||||||||||  |||||  |  |||
Db       1448 ACTCTTTGGATGTGATAGCATTCCTGCTGGCCTGGGTGGCAACTGTGATATTTATCATCA 1507

Qy       1546 CAAAATGTTTTTTATTTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGG 1605
              |||||| ||   | ||||  |  | |||  ||   ||||   ||||||   | || || |
Db       1508 CAAAATTTTGCCTGTTTTGTTTCCGAAAGCTTGCCAAAAAAGGAAAGAAGAAGAAAAGAG 1567

Qy       1606 AATAGATCTTTCCAAA 1621
              | ||| | | || |||
Db       1568 ATTAGTTATATCAAAA 1583


RESULT 11
AAZ95199
ID   AAZ95199 standard; DNA; 2092 BP.
XX
AC   AAZ95199;
XX
DT   05-JUN-2000  (first entry)
XX
DE   Human UDP-glucuronosyltransferase 2B4 nucleotide sequence.
XX
KW   UDP-glucuronosyltransferase 2B4; UGT2B4; polymorphism; metabolism; SNPs;
KW   drug interaction; detect; human; single nucleotide polymorphism; ds.
XX
OS   Homo sapiens.
XX
PN   WO200006776-A1.
XX
PD   10-FEB-2000.
XX
PF   22-JUL-1999;   99WO-US16675.
XX
PR   28-JUL-1998;   98US-0094391.
XX
PA   (AXYS-) AXYS PHARM INC.
XX
PI   Galvin M,  Miller A,  Penny L,  Riedy M;
XX
DR   WPI; 2000-195321/17.
DR   P-PSDB; AAY78933.
XX
PT   Novel human UDP-glucuronosyltransferase sequence, polymorphisms for
PT   genotyping individuals to predict rate of metabolism of substrates and
PT   for identifying potential drug interactions -
XX
PS   Disclosure; Page 34-36; 72pp; English.
XX
CC   This sequence represents the human UDP-glucuronosyltransferase 2B4
CC   (UGT2B4) gene. UDP-glucuronosyltransferase (UGTs) are a family of
CC   enzymes that catalyse the glucuronic acid conjugation of a wide range of
CC   endogenous and exogenous substrates. The UGT2B gene subfamily encode
CC   steroid metabolizing isoforms in the liver. Alteration of the expression
CC   or function of UGTs may effect drug metabolism. The invention relates to
CC   non-chromosomal nucleic acid molecules, which comprise human UGT2B
CC   sequence polymorphisms (see AAZ95051-Z95110). Probes which detect the
CC   UGT2B locus polymorphisms can be used to detect altered UGT2B metabolism
CC   of a substrate in an individual. The nucleic acid molecules comprising a
CC   human UGT2B sequence polymorphism can be used in screening assays for
CC   genotyping individuals, also to predict their rate of metabolism of
CC   UGT2B substrate, potential drug-drug interactions and adverse side
CC   effects. The polymorphisms can be used as single nucleotide polymorphisms
CC   (SNPs) for detecting genetic linkage related to phenotypic variation in
CC   activity or expression of UGT2B protein. The polymorphism containing
CC   nucleic acid molecules may also be used for generating genetically
CC   modified non-human animals and for obtaining site specific gene
CC   modification in cell lines.
XX
SQ   Sequence 2092 BP; 639 A; 398 C; 438 G; 617 T; 0 other;

  Query Match             46.8%;  Score 766;  DB 21;  Length 2092;
  Best Local Similarity   68.7%;  Pred. No. 5.5e-209;
  Matches 1102;  Conservative    0;  Mismatches  490;  Indels   12;  Gaps    3;

Qy         23 CATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGCTCCTGCAGCTCTTCTGT---GT 79
              ||||| || ||||   || |   |||  |   |||||||  | |||||   ||||    |
Db         31 CATCAGGATGTCTATGAAATGGACTTCAGCTCTTCTGCTGATACAGCTGAGCTGTTACTT 90

Qy         80 TGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTGACATGAGCCATTGGCTTAA 139
              | ||| |||    ||||| || || ||||||||||||   ||  | ||||| ||| | ||
Db         91 TAGCTCTGGGAGTTGTGGAAAGGTGCTGGTGTGGCCCACAGAATTCAGCCACTGGATGAA 150

Qy        140 TGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATGAGGTAACAGTATTGACTCA 199
              | | |||   || || || || ||  |   |||||| |||||||| || ||||||  |
Db        151 TATAAAGACAATCCTGGATGAACTTGTCCAGAGAGGTCATGAGGTGACTGTATTGGCATC 210

Qy        200 CTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCATTGAAATTTGAGGTGGTCCA 259
               |||    |   |   | ||   ||  |  || ||| |  | |||||||| ||    |
Db        211 TTCAGCTTCCATTTCTTTCGATCCCAACAGCCCATCTACTCTTAAATTTGAAGTTTATCC 270

Qy        260 TATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTGACCTAGCTCTGA------A 313
              | |  |       | ||| ||     | || ||  |     | || | |  ||
Db        271 TGTATCTTTAACTAAAACTGAGTTTGAGGATATTATCAAGCAGCTGGTTAAGAGATGGGC 330

Qy        314 TGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATGATTTTTTTGTTGA 373
               |   | |||        ||   ||| ||   ||  | || || | ||  |  |
Db        331 AGAACTTCCAAAAGACACATTTTGGTCATATTTTTCACAAGTACAAGAAATCATGTGGAC 390

Qy        374 AATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATCAGACGCTTATGAA 433
              | | |  |  |   | | || | | ||| ||    |  | |  ||| |||  ||||||||
Db        391 ATTTAATGACATACTTAGAAAGTTCTGTAAGGATATAGTTTCAAATAAGAAACTTATGAA 450

Qy        434 GAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGATTCCCTGTGGAGA 493
              ||| ||||||||  | |  |  |||||  | |||  ||| ||||  | |||| ||| ||
Db        451 GAAACTACAGGAGTCAAGATTTGATGTTGTTCTTGCAGATGCTGTTTTCCCCTTTGGTGA 510
```

```
Qy        494 CCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAATTTCTGTAGGAGG 553
Db        511 GCTGCTGGCCGAGTTACTTAAAATACCCTTTGTCTACAGCCTCCGCTTCTCTCCTGGCTA 570

Qy        554 CAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATGTACCTGTGCCTAT 613
Db        571 CGCAATTGAAAAGCATAGTGGAGGACTTCTGTTCCCTCCTTCCTATGTGCCTGTTGTTAT 630

Qy        614 GACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATTCAATGCTTTCAGT 673
Db        631 GTCAGAACTAAGTGACCAAATGACTTTCATAGAGAGGGTAAAAAATATGATCTATGTGCT 690

Qy        674 TTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAGAGTTTTATAGTAA 733
Db        691 TTATTTTGAATTTTGGTTCCAAATATTTGACATGAAGAAGTGGGATCAGTTCTACAGTGA 750

Qy        734 GGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTGAGATATGGCTAAT 793
Db        751 AGTTCTAGGAAGACCCACTACGTTATCTGAGACAATGGCAAAAGCTGACATATGGCTTAT 810

Qy        794 ACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACTTTGAGTTTGTTGG 853
Db        811 TCGAAACTACTGGGATTTTCAATTTCCTCACCCACTCTTACCAAATGTTGAGTTCGTTGG 870

Qy        854 AGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAAATTTTGTCCAGAG 913
Db        871 AGGACTCCACTGCAAACCTGCCAAACCCCTACCGAAGGAAATGGAAGAGTTTGTCCAGAG 930

Qy        914 TTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTCAAAATGTTACAGA 973
Db        931 CTCTGGAGAAAATGGTGTTGTGGTGTTTTCTCTGGGGTCGATGGTCAGTAACACGTCAGA 990

Qy        974 AGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGAAGGTGTTATGGAG 1033
Db        991 AGAAAGGGCCAATGTAATTGCATCAGCCCTTGCCAAGATCCCACAAAAGGTTCTGTGGAG 1050

Qy       1034 GTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGTATGATTGGATACC 1093
Db       1051 ATTTGATGGGAATAAACCAGATACTTTAGGACTCAATACTCGGCTGTACAAGTGGATACC 1110

Qy       1094 CCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTCATGGTGGAATGAA 1153
Db       1111 CCAGAATGATCTTCTTGGTCACCCAAAAACCAGAGCTTTTATAACTCATGGTGGAGCCAA 1170

Qy       1154 TGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTCCCATATTTGGTGA 1213
Db       1171 TGGCATCTATGAGGCAATCTACCATGGAATCCCTATGGTGGGCGTTCCATTGTTTGCAGA 1230

Qy       1214 TCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAGAAATAAACTTCAA 1273
Db       1231 TCAACCTGATAACATTGCACACATGAAGGCCAAGGGAGCAGCTGTTAGTTTGGACTTCCA 1290

Qy       1274 AACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTACCGATTCCTCTTA 1333
Db       1291 CACAATGTCGAGTACAGACTTACTCAATGCACTGAAGACAGTAATTAATGATCCTTTATA 1350

Qy       1334 TAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTGTAAAGCCCCTAGA 1393
Db       1351 TAAAGAGAATGCTATGAAATTATCAAGAATTCATCATGATCAACCAGTGAAGCCCCTTGA 1410

Qy       1394 TCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCAAGCACCTGCGATC 1453
Db       1411 TCGAGCAGTCTTCTGGATTGAATTTGTCATGCGCCATAAAGGAGCCAAGCACCTTCGGGT 1470

Qy       1454 AGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGATTGGGTTCCTGCT 1513
Db       1471 TGCAGCCCACGACCTCACCTGGTTCCAGTACCACTCTTTGGATGTGACTGGGTTCCTGCT 1530

Qy       1514 GACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTATTTTCCTGTCAAAA 1573
Db       1531 GGCCTGTGTGGCAACTGTGATATTCATCATCACAAAATGTCTGTTTTGTGTCTG---GAA 1587

Qy       1574 ATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTC 1617
Db       1588 GTTTGTTAGAACAGGAAAGAAGGGGAAAAGAGATTAATTACGTC 1631
```

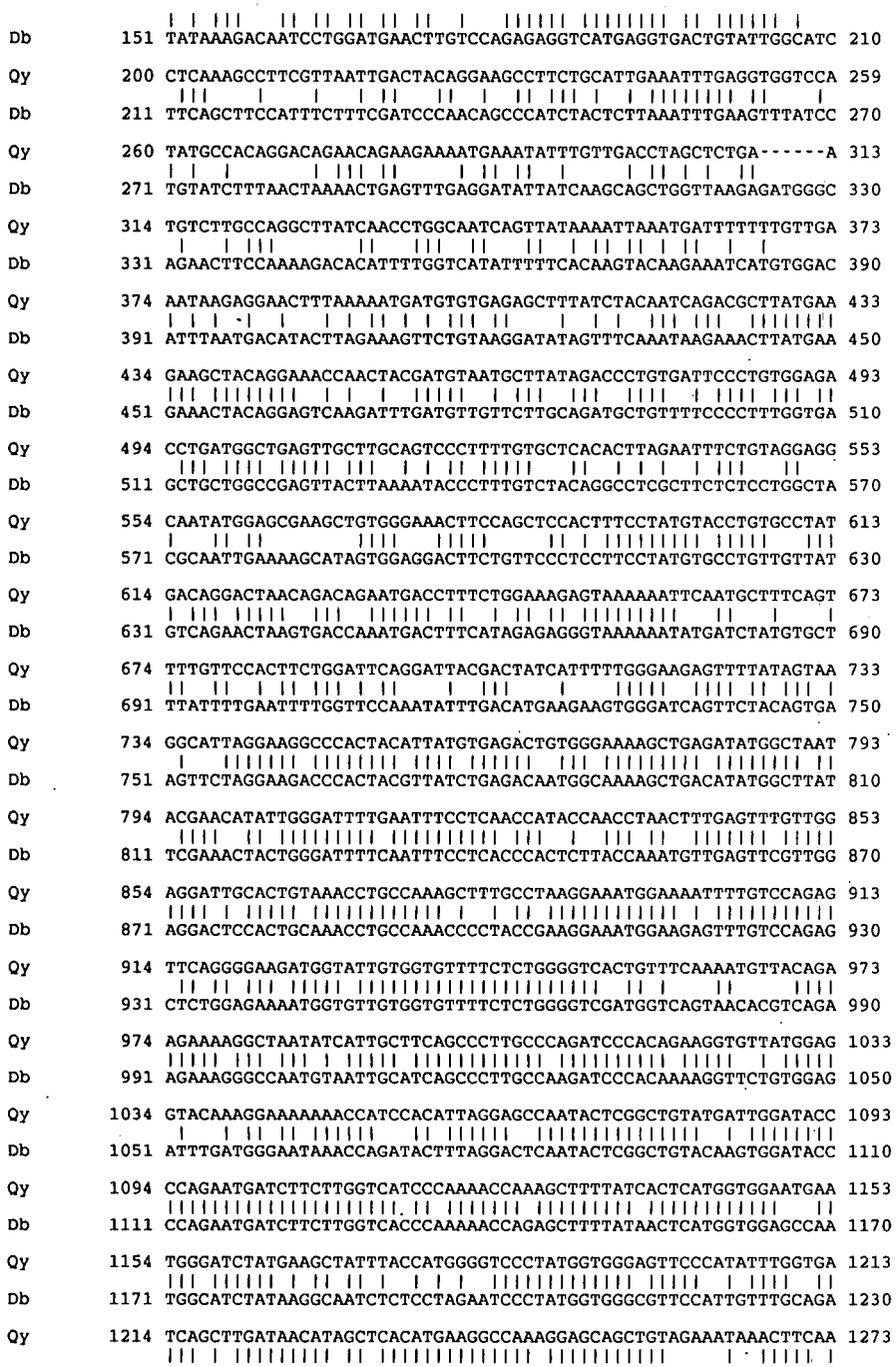
RESULT 12
ABN97347
ID ABN97347 standard; DNA; 2093 BP.
XX
AC ABN97347;
XX
DT 13-AUG-2002 (first entry)
XX
DE Gene #3845 used to diagnose liver cancer.
XX
KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW metastatic liver tumour; cytostatic; expression profile; disease state;
KW disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
OS Homo sapiens.
XX
PN WO200229103-A2.
XX
PD 11-APR-2002.
XX
PF 02-OCT-2001; 2001WO-US30589.
XX
PR 02-OCT-2000; 2000US-237054P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX
DR WPI; 2002-426119/45.
XX
PT Diagnosing and detecting the progression of liver cancer,
PT hepatocellular carcinoma or metastatic liver tumor in a patient,
PT involves detecting the level of expression of two or more genes in a
PT liver tissue sample -
XX
PS Claim 1; SEQ ID NO 3845; 298pp; English.
XX
CC The invention relates to a novel method for diagnosing and detecting the
CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC tumour in a patient, and differentiating metastatic liver cancer from
CC hepatocellular carcinoma in a patient, involving detecting the level of
CC expression of two or more genes represented in ABN93503-ABN97455 in a
CC tissue sample. The method of the invention has hepatotropic, and
CC cytostatic activity. The method is useful for diagnosing and detecting
CC the progression of liver cancer, hepatocellular carcinoma and metastatic
CC liver carcinoma in a patient. The method is useful for identifying
CC expression profiles which serve as useful diagnostic markers as well as
CC markers that can be used to monitor disease states, disease progression,
CC drug toxicity, drug efficacy and drug metabolism.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2093 BP; 638 A; 397 C; 439 G; 619 T; 0 other;

Query Match 46.3%; Score 758; DB 24; Length 2093;
Best Local Similarity 68.4%; Pred. No. 1.1e-206;
Matches 1097; Conservative 0; Mismatches 495; Indels 12; Gaps 3;

```
Qy         23 CATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGCTCCTGCAGCTCTTCTGT---GT 79
Db         31 CATCAGGATGTCTATGAAATGGACTTCAGCTCTTCTGCTGATACAGCTGAGCTGTTACTT 90

Qy         80 TGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTGACATGAGCCATTGGCTTAA 139
Db         91 TAGCTCTGGGAGTTGTGGAAAGGTGCTGGTGTGGCCCACAGAATTCAGCCACTGGATGAA 150

Qy        140 TGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATGAGGTAACAGTATTGACTCA 199
```



RESULT 13
AAD06860
ID AAD06860 standard; DNA; 1722 BP.
XX
AC AAD06860;
XX
DT 06-AUG-2001 (first entry)
XX
DE Human breast cancer specific gene-2 (BCSG-2) #2.
XX
KW Human; breast cancer specific gene-2; BCSG-2; cytostatic; vaccine;
KW breast cancer; therapeutic; gene therapy; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 10..1599
FT /*tag= a
FT /product= "Human BCSG-2 protein"
XX
PN WO200137779-A2.
XX
PD 31-MAY-2001.
XX
PF 22-NOV-2000; 2000WO-US32056.
XX
PR 23-NOV-1999; 99US-0166973.
XX
PA (DIAD-) DIADEXUS INC.
XX
PI Salceda S, Cafferkey R, Recipon H, Sun Y;
XX
DR WPI; 2001-367602/38.
DR P-PSDB; AAE02188.
XX
PT Novel breast cancer specific gene for diagnosing, monitoring, staging,
PT imaging, preventing and treating cancers, particularly breast cancer -
XX
PS Claim 1; Page 54-55; 66pp; English.
XX
CC The invention relates human breast cancer specific genes (BCSG's) and
CC their corresponding proteins. BCSG is useful for diagnosing, staging,
CC monitoring, imaging, preventing and treating breast cancers. BCSG is also
CC useful for inducing an immune response against a target cell expressing
CC BCSG. The invention also provide methods for detecting genetic lesions or
CC mutations in BCSG, thereby determining if a human with the genetic lesion

```
CC   is at risk for breast cancer or has breast cancer. BCSG antibodies
CC   labelled with paramagnetic ions or radioisotopes is useful for imaging
CC   breast cancers, while BCSG antibodies conjugated to a cytotoxic agent is
CC   useful for treating breast cancer. BCSG is useful in the rational design
CC   of new therapeutics for imaging and treating cancers. BCSG is also used
CC   in gene therapy. The present DNA sequence is human breast cancer specific
CC   gene-2 (BCSG-2) or Gene ID 480489.
XX
SQ   Sequence 1722 BP; 523 A; 319 C; 371 G; 509 T; 0 other;

  Query Match             46.1%;  Score 754.8;  DB 22;  Length 1722;
  Best Local Similarity   67.8%;  Pred. No. 8.1e-206;
  Matches 1088;  Conservative    0;  Mismatches  507;  Indels    9;  Gaps    2;

Qy         23 CATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGCTCCTGCAGCTCTTCTGT---GT  79
Db          3 CACCAGGATGACTCTGAAATGGACTTCAGTTCTTCTGCTGATACATCTCAGTTGTTACTT  62

Qy         80 TGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTGACATGAGCCATTGGCTTAA 139
Db         63 TAGCTCTGGGAGTTGTGGAAAAGTGCTGGTGTGGGCCGCAGAATACAGCCATTGGATGAA 122

Qy        140 TGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATGAGGTAACAGTATTGACTCA 199
Db        123 TATGAAGACAATCCTGAAAGAGCTTGTTCAGAGAGGTCATGAGGTGACTGTACTGGCATC 182

Qy        200 CTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCATTGAAATTTGAGGTGGTCCA 259
Db        183 TTCAGCTTCCATTCTTTTTGATCCCAATGATGCATCCACTCTTAAATTTGAAGTTTATCC 242

Qy        260 TATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTGACCTAGCTCTGA---:---A 313
Db        243 TACATCTTTAACTAAAACTGAATTTGAGAATATCATCATGCAACAGGTTAAGAGATGGTC 302

Qy        314 TGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATGATTTTTTTGTTGA 373
Db        303 AGACATTCGAAAAGATAGCTTTTGGTTATATTTTTCACAAGAACAAGAAATCCTGTGGGA 362

Qy        374 AATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATCAGACGCTTATGAA 433
Db        363 ATTATATGACATATTTAGAAACTTCTGTAAAGATGTAGTTTCAAATAAGAAAGTTATGAA 422

Qy        434 GAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGATTCCCTGTGGAGA 493
Db        423 AAAACTACAAGAGTCAAGATTTGACATCGTTTTTGCAGATGCTGTTTTTCCCTGTGGTGA 482

Qy        494 CCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAATTTCTGTAGGAGG 553
Db        483 GCTGCTGGCTGCGCTACTTAACATACGGTTTGTGTACAGTCTCCGCTTTACTCCTGGCTA 542

Qy        554 CAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATGTACCTGTGCCTAT 613
Db        543 CACAATTGAAAGGCACAGTGGAGGACTGATTTTCCCTCCTTCCTACATACCTATTGTTAT 602

Qy        614 GACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATTCAATGCTTTCAGT 673
Db        603 GTCAAAATTAAGTGATCAAATGACTTTCATGGAGAGGGTAAAAAATATGATCTATGTGCT 662

Qy        674 TTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAGAGTTTTATAGTAA 733
Db        663 TTATTTTGACTTTTGGTTCCAAATGTCTGATATGAAGAAGTGGGATCAGTTTTACAGTGA 722

Qy        734 GGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTGAGATATGGCTAAT 793
Db        723 AGTTTTAGGAAGACCCACTACCTTATTTGAGACAATGGGAAAAGCTGACATATGGCTTAT 782

Qy        794 ACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACTTTGAGTTTGTTGG 853
Db        783 GCGAAACTCCTGGAGTTTTCAATTTCCTCATCCATTCTTACCAAACGTTGATTTTGTTGG 842

Qy        854 AGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAAATTTTGTCCAGAG 913
Db        843 AGGATTCCACTGCAAACCTGCCAAACCCCTACCTAAGGAAATGGAGGAGTTTGTACAGAG 902

Qy        914 TTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTCAAAATGTTACAGA 973
Db        903 CTCTGGAGAAAATGGTGTTGTGGTGTTTTCTCTGGGGTCAGTGATAAGTAACATGACAGC 962

Qy        974 AGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGAAGGTGTTATGGAG 1033
Db        963 AGAAAGGGCCAATGTAATTGCAACAGCCCTTGCCAAGATCCCACAAAAGGTTCTGTGGAG 1022

Qy       1034 GTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGTATGATTGGATACC 1093
Db       1023 ATTTGACGGGAATAAACCAGATGCCTTAGGTCTCAATACTCGGCTGTACAAGTGGATACC 1082

Qy       1094 CCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTCATGGTGGAATGAA 1153
Db       1083 CCAGAATGACCTTCTAGGTCATCCAAAAACCAGAGCTTTTATAACTCATGGTGGAGCCAA 1142

Qy       1154 TGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTCCCATATTTGGTGA 1213
Db       1143 TGGCATCTATGAGGCAATCTACCATGGGATCCCTATGGTGGGCATTCCATTGTTTTTTGA 1202

Qy       1214 TCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAGAAATAAACTTCAA 1273
Db       1203 TCAACCTGATAACATTGCTCACATGAAGGCCAAGGGAGCAGCTGTTAGATTGGACTTCAA 1262

Qy       1274 AACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTACCGATTCCTCTTA 1333
Db       1263 CACAATGTCGAGTACAGACCTGCTGAATGCACTGAAGACAGTAATTAATGATCCTTTATA 1322

Qy       1334 TAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTGTAAAGCCCCTAGA 1393
Db       1323 TAAAGAGAATATTATGAAATTATCAAGAATTCAACATGATCAACCAGTAAAGCCCCTGGA 1382

Qy       1394 TCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCAAGCACCTGCGATC 1453
Db       1383 TCGAGCAGTCTTCTGGATTGAATTTGTCATGCCCCACAAAGGAGCCAAACACCTTCGAGT 1442

Qy       1454 AGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGATTGGGTTCCTGCT 1513
Db       1443 TGCAGCCCATGACCTCACCTGGTTCCAGTACCACTCTTTGGATGTGATTGGGTTTCTGCT 1502

Qy       1514 GACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTATTTTCCTGTCAAAA 1573
Db       1503 GGCCTGTGTGGCAACTGTGATATTTATCATCACAAAGTTTTGTCTGTTTTGTTTCTGGAA 1562

Qy       1574 ATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTC 1617
Db       1563 GTTTGCTAGAAAAGGGAAGAAGGGAAAAAGAGATTAGTTATGTC 1606

RESULT 14
ABN95622
ID   ABN95622 standard; DNA; 1855 BP.
XX
AC   ABN95622;
XX
DT   13-AUG-2002  (first entry)
XX
DE   Gene #2120 used to diagnose liver cancer.
XX
KW   Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW   metastatic liver tumour; cytostatic; expression profile; disease state;
KW   disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
OS   Homo sapiens.
XX
PN   WO200229103-A2.
XX
PD   11-APR-2002.
XX
PF   02-OCT-2001; 2001WO-US30589.
XX
PR   02-OCT-2000; 2000US-237054P.
```

XX
PA (GENE-) GENE LOGIC INC.
XX
PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX
DR WPI; 2002-426119/45.
XX
PT Diagnosing and detecting the progression of liver cancer,
PT hepatocellular carcinoma or metastatic liver tumor in a patient,
PT involves detecting the level of expression of two or more genes in a
PT liver tissue sample -
XX
PS Claim 1; SEQ ID NO 2120; 298pp; English.
XX
CC The invention relates to a novel method for diagnosing and detecting the
CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC tumour in a patient, and differentiating metastatic liver cancer from
CC hepatocellular carcinoma in a patient, involving detecting the level of
CC expression of two or more genes represented in ABN93503-ABN97455 in a
CC tissue sample. The method of the invention has hepatotropic, and
CC cytostatic activity. The method is useful for diagnosing and detecting
CC the progression of liver cancer, hepatocellular carcinoma and metastatic
CC liver carcinoma in a patient. The method is useful for identifying
CC expression profiles which serve as useful diagnostic markers as well as
CC markers that can be used to monitor disease states, disease progression,
CC drug toxicity, drug efficacy and drug metabolism.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1855 BP; 571 A; 337 C; 394 G; 553 T; 0 other;

Query Match 46.1%; Score 754.8; DB 24; Length 1855;
Best Local Similarity 67.8%; Pred. No. 8.5e-206;
Matches 1088; Conservative 0; Mismatches 507; Indels 9; Gaps 2;

```
Qy        23 CATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGCTCCTGCAGCT---CTTCTGTGT 79
             || || || |||||  || |   |||  ||| || ||||  | || ||   ||| ||  |
Db         8 CACCAGGATGTCTGTGAAATGGACTTCAGTAATTTTGCTAATACAACTGAGCTTTTGCTT 67

Qy        80 TGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTGACATGAGCCATTGGCTTAA 139
             | ||| |||    ||||| || || ||||||||| |    ||    ||||||||| | ||
Db        68 TAGCTCTGGGAATTGTGGAAAGGTGCTGGTGTGGGCAGCAGAATACAGCCATTGGATGAA 127

Qy       140 TGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATGAGGTAACAGTATTGACTCA 199
             | | |||   || || || ||||| ||   |||||| |||||||| || ||| || |
Db       128 TATAAAGACAATCCTGGATGAGCTTATTCAGAGAGGTCATGAGGTGACTGTACTGGCATC 187

Qy       200 CTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCATTGAAATTTGAGGTGGTCCA 259
             .|||    |     |  ||||   ||  ||  | || ||  | ||| ||||  |    |
Db       188 TTCAGCTTCCATTCTTTTGATCCCAACAACTCATCCGCTCTTAAAATTGAAATTTATCC 247

Qy       260 TATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTGACCTAGCTCTGA------A 313
             |   |       | ||| ||     |  |  |  |  |  | |    |  ||
Db       248 CACATCTTTAACTAAAACTGAGTTGGAGAATTTCATCATGCAACAGATTAAGAGATGGTC 307

Qy       314 TGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATGATTTTTTTGTTGA 373
             | | | |||       ||   |||  ||   ||  | || || | ||  |  |
Db       308 AGACCTTCCAAAAGATACATTTTGGTTATATTTTTCACAAGTACAGGAAATCATGTCAAT 367

Qy       374 AATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATCAGACGCTTATGAA 433
             | |  | |  |     | || | | ||| |     |  | |  ||| |||   |||||||
Db       368 ATTTGGTGACATAACTAGAAAGTTCTGTAAAGATGTAGTTTCAAATAAGAAATTTATGAA 427

Qy       434 GAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGATTCCCTGTGGAGA 493
             ||  |||| ||   | |  |  || || ||  ||  |||  || |  |||||||| | ||
Db       428 AAAAGTACAAGAGTCAAGATTTGACGTCATTTTTGCAGATGCTATTTTTCCCTGTAGTGA 487

Qy       494 CCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAATTTCTGTAGGAGG 553
             |||  |||||||| |  ||    | || ||||||  ||  || ||  | |||   ||
Db       488 GCTGCTGGCTGAGCTATTTAACATACCCTTTGTGTACAGTCTCAGCTTCTCTCCTGGCTA 547

Qy       554 CAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATGTACCTGTGCCTAT 613
                || | | ||        ||||   | ||      || | |||||| ||||||||   |||
Db       548 CACTTTTGAAAAGCATAGTGGAGGATTTATTTTCCCTCCTTCCTACGTACCTGTTGTTAT 607

Qy       614 GACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATTCAATGCTTTCAGT 673
             | ||| | |||| ||   |||||| ||  |||| || |||||||||   ||   |    |
Db       608 GTCAGAATTAACTGATCAAATGACTTTCATGGAGAGGGTAAAAAATATGATCTATGTGCT 667

Qy       674 TTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAGAGTTTTATAGTAA 733
             ||  ||  |||| ||| |  |    |  |||    |    ||||| |||||||||| |
Db       668 TTACTTTGACTTTTGGTTCGAAATATTTGACATGAAGAAGTGGGATCAGTTTTATAGTGA 727

Qy       734 GGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTGAGATATGGCTAAT 793
             |   ||||||| |||||||| |||| ||||||  |||| |||||||| |||||||  ||
Db       728 AGTTCTAGGAAGACCCACTACGTTATCTGAGACAATGGGGAAAGCTGACGTATGGCTTAT 787

Qy       794 ACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACTTTGAGTTTGTTGG 853
             ||||  |  ||| ||||| | |||||||| |||  |  ||| ||  |||| |||||||
Db       788 TCGAAACTCCTGGAATTTTCAGTTTCCTCATCCACTCTTACCAAATGTTGATTTTGTTGG 847

Qy       854 AGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAAATTTTGTCCAGAG 913
             |||| | ||||| |||||||||||| |  |||||||||||||||| | ||||| |||||
Db       848 AGGACTCCACTGCAAACCTGCCAAACCCCTGCCTAAGGAAATGGAAGACTTTGTACAGAG 907

Qy       914 TTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTCAAAATGTTACAGA 973
             || || ||| |||| ||||||||||||||||||||||||| || |    ||  | ||||
Db       908 CTCTGGAGAAAATGGTGTTGTGGTGTTTTCTCTGGGGTCAATGGTCAGTAACATGACAGA 967

Qy       974 AGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGAAGGTGTTATGGAG 1033
             ||||| ||| ||  | ||||| |||||||| |||||||||||||| |||||  | |||||
Db       968 AGAAAGGGCCAACGTAATTGCATCAGCCCTGGCCCAGATCCCACAAAAGGTTCTGTGGAG 1027

Qy      1034 GTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGTATGATTGGATACC 1093
             |   | || || ||||||   || |||||   |||||||||||||||| | ||||||||
Db      1028 ATTTGATGGGAATAAACCAGATACCTTAGGTCTCAATACTCGGCTGTATAAGTGGATACC 1087

Qy      1094 CCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTCATGGTGGAATGAA 1153
             |||||||| ||||| |||||||| || |||| |||||||| ||||||||||    ||
Db      1088 CCAGAATGACCTTCTAGGTCATCCAAAGACCAGAGCTTTTATAACTCATGGTGGAGCCAA 1147

Qy      1154 TGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTCCCATATTTGGTGA 1213
             ||| ||||| || || || ||||||||| ||||||||||||  ||||  | ||||  ||
Db      1148 TGGCATCTACGAGGCAATCTACCATGGGATCCCTATGGTGGGGATTCCATTGTTTGCCGA 1207

Qy      1214 TCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAGAAATAAACTTCAA 1273
             ||| | |||||||||| ||||||||||||||||  ||||||||||   | |  |||||||
Db      1208 TCAACCTGATAACATTGCTCACATGAAGGCCAGGGGAGCAGCTGTTAGAGTGGACTTCAA 1267

Qy      1274 AACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTACCGATTCCTCTTA 1333
             || ||| | ||   ||| || |||| || ||||  | ||| ||||  ||| | || ||
Db      1268 CACAATGTCGAGTACAGACTTGCTGAATGCATTGAAGAGAGTAATTAATGATCCTTCATA 1327

Qy      1334 TAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTGTAAAGCCCCTAGA 1393
             |||||||||| ||||| |||||||||||||| |||||||||| || |||||||| ||
Db      1328 TAAAGAGAATGTTATGAAATTATCAAGAATTCAACATGATCAACCAGTGAAGCCCCTGGA 1387

Qy      1394 TCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCAAGCACCTGCGATC 1453
             |||||||||||||||||| || |||||||||||||||||||||| || ||||| ||
Db      1388 TCGAGCAGTCTTCTGGATTGAATTTGTCATGCGCCACAAAGGAGCTAAACACCTTCGGGT 1447

Qy      1454 AGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGATTGGGTTCCTGCT 1513
             || ||||| |||||||||||||||||| || ||||| | ||||||||||||||||||||
Db      1448 TGCAGCCCACGACCTCACCTGGTTCCAGTACCACTCTTTGGATGTGATTGGGTTCCTGCT 1507

Qy      1514 GACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTATTTTCCTGTCAAAA 1573
             |  |||||||||||||  ||||| | |||||||||||| | | ||||  |    ||
Db      1508 GGTCTGTGTGGCAACTGTGATATTTATCGTCACAAAATGTTGTCTGTTTTGTTTCTGGAA 1567

Qy      1574 ATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTC 1617
             |||  || ||    |||||  | ||| |  || ||| | | ||
Db      1568 GTTTGCTAGAAAAGCAAAGAAGGGAAAAAATGATTAGTTATATC 1611
```

RESULT 15

ABL68501
ID ABL68501 standard; DNA; 1855 BP.
XX
AC ABL68501;
XX
DT 15-MAY-2002 (first entry)
XX
DE Kidney cancer related gene sequence SEQ ID NO:6838.
XX
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200194629-A2.
XX
PD 13-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US10838.
XX
PR 05-JUN-2000; 2000US-209473P.
PR 05-JUN-2000; 2000US-209531P.
PR 18-SEP-2000; 2000US-233133P.
PR 18-SEP-2000; 2000US-233617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 20-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.
PR 25-SEP-2000; 2000US-234923P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-235077P.
PR 25-SEP-2000; 2000US-235082P.
PR 25-SEP-2000; 2000US-235134P.
PR 25-SEP-2000; 2000US-235280P.
PR 26-SEP-2000; 2000US-235637P.
PR 26-SEP-2000; 2000US-235638P.
PR 27-SEP-2000; 2000US-235711P.
PR 27-SEP-2000; 2000US-235720P.
PR 27-SEP-2000; 2000US-235840P.
PR 27-SEP-2000; 2000US-235863P.
PR 28-SEP-2000; 2000US-236028P.
PR 28-SEP-2000; 2000US-236032P.
PR 28-SEP-2000; 2000US-236033P.
PR 28-SEP-2000; 2000US-236034P.
PR 28-SEP-2000; 2000US-236109P.
PR 28-SEP-2000; 2000US-236111P.
PR 29-SEP-2000; 2000US-236842P.
PR 29-SEP-2000; 2000US-236891P.
PR 02-OCT-2000; 2000US-237172P.
PR 02-OCT-2000; 2000US-237173P.
PR 02-OCT-2000; 2000US-237278P.
PR 02-OCT-2000; 2000US-237294P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237316P.
PR 03-OCT-2000; 2000US-237425P.
PR 03-OCT-2000; 2000US-237598P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.
PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.
XX
PA (AVAL-) AVALON PHARM.
XX
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX
DR WPI; 2002-188264/24.
XX
PT Screening for anti-neoplastic agent involves exposing cells to a
PT chemical agent to be tested for anti-neoplastic activity, and
PT determining a change in expression of a gene of a signature gene set -
XX
PS Claim 1; SEQ ID 6838; 44pp; English.
XX
CC The present invention describes a method (M1) for screening for an
CC anti-neoplastic agent. The method involves exposing cells to a chemical
CC agent to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening
CC an anti-neoplastic agent, and can be used for producing a product which
CC is the data collected with respect to the anti-neoplastic agent as a
CC result of M1, and the data is sufficient to convey the chemical
CC structure and/or properties of the agent. M1 can be used in the
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC carcinoma, papillary carcinoma and Wilm's tumour.
XX
SQ Sequence 1855 BP; 571 A; 337 C; 394 G; 553 T; 0 other;

Query Match 46.1%; Score 754.8; DB 24; Length 1855;
Best Local Similarity 67.8%; Pred. No. 8.5e-206;
Matches 1088; Conservative 0; Mismatches 507; Indels 9; Gaps 2;

```
Qy      23 CATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGCTCCTGCAGCT---CTTCTGTGT 79
           || || || |||||  || |   |||  ||| || ||||  | || ||   ||| ||  |
Db       8 CACCAGGATGTCTGTGAAATGGACTTCAGTAATTTTGCTAATACAACTGAGCTTTTGCTT 67

Qy      80 TGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTGACATGAGCCATTGGCTTAA 139
           | ||| |||    ||||| || || |||||||||  |    ||    |||||||| | ||
Db      68 TAGCTCTGGGAATTGTGGAAAGGTGCTGGTGTGGGCAGCAGAATACAGCCATTGGATGAA 127

Qy     140 TGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATGAGGTAACAGTATTGACTCA 199
           | | |||   || || || ||||| ||   |||||| |||||||| || ||| || |
Db     128 TATAAAGACAATCCTGGATGAGCTTATTCAGAGAGGTCATGAGGTGACTGTACTGGCATC 187

Qy     200 CTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCATTGAAATTTGAGGTGGTCCA 259
           |||    |    |  ||||   ||  ||  | || ||  | ||| ||||  |    |
Db     188 TTCAGCTTCCATTCTTTTTGATCCCAACAACTCATCCGCTCTTAAAATTGAAATTTATCC 247

Qy     260 TATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTGACCTAGCTCTGA------A 313
           |   |       | ||| ||     |  |  |  |  |  |  |  |    | ||
Db     248 CACATCTTTAACTAAAACTGAGTTGGAGAATTTCATCATGCAACAGATTAAGAGATGGTC 307

Qy     314 TGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATGATTTTTTTGTTGA 373
           | | | |||        ||   |||  ||   ||  | || || | ||  |  |
Db     308 AGACCTTCCAAAAGATACATTTTGGTTATATTTTTCACAAGTACAGGAAATCATGTCAAT 367

Qy     374 AATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATCAGACGCTTATGAA 433
           | |  | |  |    | || | | ||| |     |  | |  ||| |||   |||||||
Db     368 ATTTGGTGACATAACTAGAAAGTTCTGTAAAGATGTAGTTTCAAATAAGAAATTTATGAA 427

Qy     434 GAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGATTCCCTGTGGAGA 493
           ||  |||| ||  | |  |  || || ||  ||  |||  || |  |||||||| | ||
Db     428 AAAAGTACAAGAGTCAAGATTTGACGTCATTTTTGCAGATGCTATTTTTCCCTGTAGTGA 487

Qy     494 CCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAATTTCTGTAGGAGG 553
           |||  |||||||| |  ||    | || ||||||  ||  ||  ||  | |||   ||
Db     488 GCTGCTGGCTGAGCTATTTAACATACCCTTTGTGTACAGTCTCAGCTTCTCTCCTGGCTA 547

Qy     554 CAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATGTACCTGTGCCTAT 613
           || | | ||          ||||   | ||      || | |||||| ||||||||   |||
Db     548 CACTTTTGAAAAGCATAGTGGAGGATTTATTTTCCCTCCTTCCTACGTACCTGTTGTTAT 607

Qy     614 GACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATTCAATGCTTTCAGT 673
           | ||| | |||| ||   |||||| ||  |||| || |||||||||   ||   |    |
Db     608 GTCAGAATTAACTGATCAAATGACTTTCATGGAGAGGGTAAAAAATATGATCTATGTGCT 667
```

```
Qy       674 TTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAGAGTTTTATAGTAA 733
Db       668 TTACTTTGACTTTTGGTTCGAAATATTTGACATGAAGAAGTGGGATCAGTTTTATAGTGA 727

Qy       734 GGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTGAGATATGGCTAAT 793
Db       728 AGTTCTAGGAAGACCCACTACGTTATCTGAGACAATGGGGAAAGCTGACGTATGGCTTAT 787

Qy       794 ACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACTTTGAGTTTGTTGG 853
Db       788 TCGAAACTCCTGGAATTTTCAGTTTCCTCATCCACTCTTACCAAATGTTGATTTTGTTGG 847

Qy       854 AGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAAATTTTGTCCAGAG 913
Db       848 AGGACTCCACTGCAAACCTGCCAAACCCCTGCCTAAGGAAATGGAAGACTTTGTACAGAG 907

Qy       914 TTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTCAAAATGTTACAGA 973
Db       908 CTCTGGAGAAAATGGTGTTGTGGTGTTTTCTCTGGGGTCAATGGTCAGTAACATGACAGA 967

Qy       974 AGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGAAGGTGTTATGGAG 1033
Db       968 AGAAAGGGCCAACGTAATTGCATCAGCCCTGGCCCAGATCCCACAAAAGGTTCTGTGGAG 1027

Qy      1034 GTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGTATGATTGGATACC 1093
Db      1028 ATTTGATGGGAATAAACCAGATACCTTAGGTCTCAATACTCGGCTGTATAAGTGGATACC 1087

Qy      1094 CCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTCATGGTGGAATGAA 1153
Db      1088 CCAGAATGACCTTCTAGGTCATCCAAAGACCAGAGCTTTTATAACTCATGGTGGAGCCAA 1147

Qy      1154 TGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTCCCATATTTGGTGA 1213
Db      1148 TGGCATCTACGAGGCAATCTACCATGGGATCCCTATGGTGGGGATTCCATTGTTTGCCGA 1207

Qy      1214 TCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAGAAATAAACTTCAA 1273
Db      1208 TCAACCTGATAACATTGCTCACATGAAGGCCAGGGGAGCAGCTGTTAGAGTGGACTTCAA 1267

Qy      1274 AACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTACCGATTCCTCTTA 1333
Db      1268 CACAATGTCGAGTACAGACTTGCTGAATGCATTGAAGAGAGTAATTAATGATCCTTCATA 1327

Qy      1334 TAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTGTAAAGCCCCTAGA 1393
Db      1328 TAAAGAGAATGTTATGAAATTATCAAGAATTCAACATGATCAACCAGTGAAGCCCCTGGA 1387

Qy      1394 TCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCAAGCACCTGCGATC 1453
Db      1388 TCGAGCAGTCTTCTGGATTGAATTTGTCATGCGCCACAAAGGAGCTAAACACCTTCGGGT 1447

Qy      1454 AGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGATTGGGTTCCTGCT 1513
Db      1448 TGCAGCCCACGACCTCACCTGGTTCCAGTACCACTCTTTGGATGTGATTGGGTTCCTGCT 1507

Qy      1514 GACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTATTTTCCTGTCAAAA 1573
Db      1508 GGTCTGTGTGGCAACTGTGATATTTATCGTCACAAAATGTTGTCTGTTTTGTTTCTGGAA 1567

Qy      1574 ATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTC 1617
Db      1568 GTTTGCTAGAAAAGCAAAGAAGGGAAAAAATGATTAGTTATATC 1611
```

Search completed: July 16, 2003, 11:10:05
Job time : 285.458 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 16, 2003, 08:51:42 ; Search time 1786.01 Seconds
(without alignments)
14835.236 Million cell updates/sec

Title: US-09-981-353-165
Perfect score: 1636
Sequence: 1 gatcagtgtgtgagggaact.....ccaaattcaagaaagacctg 1636

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	882.8	54.0	2099	11	AK008601	AK008601 Mus muscu
2	698.8	42.7	2573	11	AK004971	AK004971 Mus muscu
3	625.2	38.2	960	14	BQ713091	BQ713091 AGENCOURT
4	623.2	38.1	1004	14	BQ942104	BQ942104 AGENCOURT
5	622.4	38.0	1896	11	AK002736	AK002736 Mus muscu
6	600	36.7	927	14	BQ925596	BQ925596 AGENCOURT
7	552.4	33.8	930	14	BQ934372	BQ934372 AGENCOURT
8	520.4	31.8	1030	14	BQ714919	BQ714919 AGENCOURT
9	492.6	30.1	913	14	BQ714368	BQ714368 AGENCOURT
c 10	477	29.2	823	12	BF689099	BF689099 602185172
11	467.4	28.6	671	12	BG249053	BG249053 602361828
12	464.4	28.4	977	13	BI331895	BI331895 602982580
13	464	28.4	871	14	BQ900474	BQ900474 AGENCOURT
14	463.4	28.3	938	14	BQ942682	BQ942682 AGENCOURT
15	456.4	27.9	891	12	BG189133	BG189133 RST8173 A
16	455.6	27.8	770	12	BG201991	BG201991 RST21337
17	435	26.6	746	12	BF688309	BF688309 602185172
18	424.8	26.0	996	14	BQ714776	BQ714776 AGENCOURT
19	418.6	25.6	801	12	BG212458	BG212458 RST32045
20	399.6	24.4	759	13	BI330877	BI330877 602981265
21	399.4	24.4	554	10	BE013295	BE013295 123199 MA
22	398	24.3	962	13	BG969620	BG969620 602836951
23	394	24.1	728	12	BG569536	BG569536 602588811
c 24	394	24.1	761	9	AI118428	AI118428 ue40e09.x
25	387.2	23.7	1005	12	BG167171	BG167171 602344696
26	383	23.4	1192	12	BF789664	BF789664 602105279
27	378.4	23.1	886	12	BG402016	BG402016 602466734
28	373	22.8	651	9	AA255327	AA255327 mz84f10.r
29	369.2	22.6	830	12	BG195984	BG195984 RST15067
c 30	367.2	22.4	836	9	AI529553	AI529553 ui82a02.x
31	366.2	22.4	935	12	BG427877	BG427877 602501564
32	364.2	22.3	812	13	BI221130	BI221130 602939634
33	361.4	22.1	823	12	BG193837	BG193837 RST12973
34	351.8	21.5	882	13	BI144244	BI144244 602907984
35	349.6	21.4	906	9	AI529750	AI529750 ui82a02.y
36	342.8	21.0	814	12	BG428781	BG428781 602500860
37	340.4	20.8	513	12	BF105272	BF105272 601822259
38	337.8	20.6	1022	14	BM925650	BM925650 AGENCOURT
39	334.2	20.4	696	13	BG971909	BG971909 602841666
40	332.6	20.3	2450	11	BC011409	BC011409 Homo sapi
41	331.4	20.3	557	12	BF080461	BF080461 231360 MA
c 42	331	20.2	639	12	BF322562	BF322562 maa28e12.
43	330.2	20.2	934	12	BF786200	BF786200 602110490
44	328.6	20.1	2202	11	AK002629	AK002629 Mus muscu
c 45	321.6	19.7	892	12	BG208263	BG208263 RST27755

ALIGNMENTS

RESULT 1
AK008601
LOCUS AK008601 2099 bp mRNA linear HTC 19-JAN-2002
DEFINITION Mus musculus adult male small intestine cDNA, RIKEN full-length enriched library, clone:2010321J07:homolog to UDP GLUCURONOSYLTRANSFERASE UGT2A3, full insert sequence.
ACCESSION AK008601
VERSION AK008601.1 GI:12842884
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male small intestine cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library clone:2010321J07.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159

```
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660
   PUBMED   11217851
REFERENCE   5  (bases 1 to 2099)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
            Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,
            Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,
            Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,
            Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,
            Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
            Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
            Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
            Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I.,
            Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
            Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .2099
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:2010321J07"
                     /db_xref="MGD:MGI:1904826"
                     /db_xref="taxon:10090"
                     /clone="2010321J07"
                     /sex="male"
                     /tissue_type="small intestine"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             15. .1619
                     /note="data source:SPTR, source key:Q9R110, evidence:ISS
                     homolog to UDP GLUCURONOSYLTRANSFERASE UGT2A3
                     putative"
                     /codon_start=1
                     /protein_id="BAB25770.1"
                     /db_xref="GI:12842885"
                     /db_xref="MGD:MGI:1919344"
                     /translation="MVSEKCVAAFFLLQLCWAGCGFCSKVLVWPCDMSHWLNLKTILE
                     ELGARGHEVTVLKYPSIIIDQSKRIPLHFENIPLLYEIETAENRLNEIANLAVNVIPN
                     LSLWEAAKTLQDFFLQVTGDFESICRSVLYNQKFMDKLRDAQYDVVVIDPVVPCGELV
                     AEVLQIPFVYTLRFSMGYYMEKHCGQLPIPLSYVPVVMSELTDNMTFTERVKNMMFSL
                     LFEYWLQQYDFAFWDQFYSETLGRPTTFCKTVGEADIWLIRTYWDVEFPRPYLPNFEF
                     VGGLHCKPAKPLPKEMEEFVQSSGEHGVVVFSLGSMVKNLTEEKANLIASVLAQIPQK
                     VLWRYSGKKPAILGSNTRLFNWIPQNDLLGHPKTKAFITHGGTNGIYEAIYHGVPMVG
                     VPMLGDQPHNIAHMEAKGAALKVSISTMTSTDLLSAVRAVINEPSYKENAMRLSRIHH
                     DQPVKPLDRAVFWIEFVMRHKGAKHLRVAAHDLSWFQYHSLDVIGFLLLCVVTLTFII
                     TKFCLFVCQKLYMKESKKMGNRKKKN"
BASE COUNT      618 a    406 c    467 g    608 t
ORIGIN
```

```
  Query Match           54.0%;  Score 882.8;  DB 11;  Length 2099;
  Best Local Similarity  73.4%;  Pred. No. 8.4e-204;
  Matches 1173;  Conservative    0;  Mismatches  412;  Indels   13;  Gaps    3;

Qy         23 CATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGCTCCTGCAGCTCTTCTGTGTTGG 82
              ||  |||   ||||| || |  | |  || |||| | || |||||||| | ||| |  ||
Db         11 CAATATGGTCTCTGAAAAATGTGTTGCGGCATTTTTTCTGCTGCAGCTTTGCTGGGCCGG 70

Qy         83 CTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTGACATGAGCCATTGGCTTAATGT 142
              ||||||||||||  | || ||||| |||||||||||||| ||||||||| ||||| | |
Db         71 CTGTGGATTCTGCAGCAAGGTCCTCGTGTGGCCCTGTGATATGAGCCACTGGCTGAATCT 130

Qy        143 CAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATGAGGTAACAGTATTGACTCACTC 202
                |||   ||||| || |||||   ||  ||||| || ||||||||||| |||   || |
Db        131 AAAGACTATTCTTGAGGAGCTTGGAGCAAGAGGGCACGAGGTAACAGTCCTGAAATACCC 190

Qy        203 AAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCATTGAAATTTGAGGTGGTCCATAT 262
              |    |    | || ||  | || || | |  | || || | |||||    | | | |
Db        191 CA---GTATCATCATAGATCAGAGTAAACGTATTCCACTGCACTTTGAGAATATTCCTTT 247

Qy        263 GCCACAGGACAGAACAGAAGAAAATGAAATAT---------TTGTTGACCTAGCTCTGAA 313
              ||   | || |        ||   |  |         |||   | |||||| ||||
Db        248 GCTGTATGAAATCGAGACAGCTGAGAATCGTTTAAATGAGATTGCAAATCTAGCTGTGAA 307

Qy        314 TGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATGATTTTTTTGTTGA 373
              |||| | |||  | | |||   ||| || |||  | || |||| | || || ||| || |
Db        308 TGTCATTCCAAACCTGTCACTGTGGGAAGCAGCAAAAACATTACAAGACTTCTTTCTTCA 367

Qy        374 AATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATCAGACGCTTATGAA 433
              | |||  |||  |||  |||  || |||  |||  |  | |||| |||   | ||| |
Db        368 AGTAACTGGAGATTTTGAAAGTATTTGTAGGAGTGTATTGTACAACCAGAAATTCATGGA 427

Qy        434 GAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGATTCCCTGTGGAGA 493
              |||||||  |||  |  | || |||||| || ||||||||||||  ||||||||||||
Db        428 CAAGCTACGGGATGCACAATATGATGTAGTGGTTATAGACCCTGTCGTTCCCTGTGGAGA 487

Qy        494 CCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAATTTCTGTAGGAGG 553
                || |||| ||  |||||    ||||||| ||   |||||| ||  |      | ||
Db        488 GTTGGTGGCAGAAGTGCTTCAGATCCCTTTCGTATACACACTGAGGTTCAGCATGGGCTA 547

Qy        554 CAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATGTACCTGTGCCTAT 613
              | | ||||||  |  |||||  | |||||  |||||| || ||||||||  ||    ||
Db        548 CTACATGGAGAAACACTGTGGCCAGCTTCCAATTCCACTCTCGTATGTACCGGTTGTCAT 607

Qy        614 GACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATTCAATGCTTTCAGT 673
              ||  |  || |||||||  ||||||||     || || || ||||||   ||| ||||| |
Db        608 GAGTGAGCTGACAGACAATATGACCTTCACAGAGAGGGTGAAAAATATGATGTTTTCACT 667

Qy        674 TTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAGAGTTTTATAGTAA 733
              |||||  | | |||| | ||| | || |||| |   || |||||  ||||||| ||| |
Db        668 GTTGTTTGAGTACTGGCTCCAGCAATATGACTTTGCATTCTGGGATCAGTTTTACAGTGA 727
```



```
Qy        734 GGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTGAGATATGGCTAAT 793
Db        728 AACCCTAGGAAGGCCCACAACGTTCTGTAAGACTGTGGGGGAAGCTGACATTTGGCTAAT 787

Qy        794 ACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACTTTGAGTTTGTTGG 853
Db        788 CCGAACATATTGGGATGTTGAGTTTCCTCGTCCATATTTACCAAATTTTGAGTTTGTGGG 847

Qy        854 AGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAAATTTTGTCCAGAG 913
Db        848 AGGACTGCACTGCAAACCTGCCAAGCCTTTACCTAAGGAAATGGAAGAATTTGTTCAGAG 907

Qy        914 TTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTCAAAATGTTACAGA 973
Db        908 CTCTGGAGAACATGGTGTAGTAGTATTTTCACTGGGGTCAATGGTCAAAAACCTGACAGA 967

Qy        974 AGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGAAGGTGTTATGGAG 1033
Db        968 AGAGAAAGCCAACCTCATTGCCTCTGTCCTTGCCCAGATTCCCCAGAAGGTTTTGTGGAG 1027

Qy       1034 GTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGTATGATTGGATACC 1093
Db       1028 ATACTCAGGCAAGAAGCCAGCCATATTAGGATCCAATACTCGGCTTTTTAATTGGATTCC 1087

Qy       1094 CCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTCATGGTGGAATGAA 1153
Db       1088 CCAGAATGATCTTCTTGGACATCCTAAAACCAAAGCTTTCATCACACATGGTGGAACAAA 1147

Qy       1154 TGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTCCCATATTTGGTGA 1213
Db       1148 CGGGATTTATGAAGCCATTTACCATGGGGTCCCTATGGTGGGCGTTCCCATGTTAGGGGA 1207

Qy       1214 TCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAGAAATAAACTTCAA 1273
Db       1208 TCAGCCTCACAACATCGCTCACATGGAGGCCAAGGGAGCAGCCCTGAAAGTCAGCATCAG 1267

Qy       1274 AACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTACCGATTCCTCTTA 1333
Db       1268 TACAATGACGAGCACAGATTTACTCAGTGCTGTGAGGGCAGTCATTAATGAGCCTTCTTA 1327

Qy       1334 TAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTGTAAAGCCCCTAGA 1393
Db       1328 TAAAGAGAATGCCATGCGGTTATCAAGAATCCACCATGATCAGCCAGTGAAGCCCCTGGA 1387

Qy       1394 TCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCAAGCACCTGCGATC 1453
Db       1388 CAGAGCAGTCTTCTGGATTGAGTTTGTCATGCGTCACAAAGGAGCCAAGCATCTTCGTGT 1447

Qy       1454 AGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGATTGGGTTCCTGCT 1513
Db       1448 GGCAGCCCATGACCTCAGCTGGTTTCAGTACCACTCCCTAGATGTGATTGGGTTCCTATT 1507

Qy       1514 GACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTATTTTCCTGTCAAAA 1573
Db       1508 GTTGTGTGTCGTTACTCTGACATTCATCATCACTAAATTTTGTTTGTTTGTGTGTCAAAA 1567

Qy       1574 ATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGA 1611
Db       1568 ACTTTATATGA-AAGAAAGTAAGAAAATGGGGAACAGA 1604
```

RESULT 2
AK004971
LOCUS AK004971 2573 bp mRNA linear HTC 19-JAN-2002
DEFINITION Mus musculus adult male liver cDNA, RIKEN full-length enriched library, clone:1300012D20:homolog to UDP-GLUCURONOSYLTRANSFERASE 2B1 PRECURSOR, MICROSOMAL (EC 2.4.1.17) (UDPGT) (UDPGTR-2), full insert sequence.
ACCESSION AK004971
VERSION AK004971.1 GI:12836559
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male liver cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library clone:1300012D20.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M.,.Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I., Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R., Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T., Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H., Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G., Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M., Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H., Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N., Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C., Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D., Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P., Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L., Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S. and Hayashizaki,Y.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409 (6821), 685-690 (2001)
MEDLINE 21085660
PUBMED 11217851
REFERENCE 5 (bases 1 to 2573)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F., Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of

Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Gemomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)

COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5′ GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTTVN 3′], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence[5′ GAGAGAGAGAAGGATCCAAGAGCTCAATTAATTTAATTAAACCCCCCCCCCC 3′]. cDNA was cleaved with XhoI and SstI. Cloning sites, 5′ end: SstI; 3′ end: XhoI. Host: SOLR.

FEATURES Location/Qualifiers
source 1. .2573
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:1300012D20"
/db_xref="MGD:MGI:1904432"
/db_xref="taxon:10090"
/clone="1300012D20"
/sex="male"
/tissue_type="liver"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
misc_feature 1. .2573
/note="data source:SPTR, source key:P09875, evidence:ISS homolog to UDP-GLUCURONOSYLTRANSFERASE 2B1 PRECURSOR, MICROSOMAL (EC 2.4.1.17) (UDPGT) (UDPGTR-2)"
/db_xref="MGD:MGI:1919023"

BASE COUNT 733 a 541 c 497 g 802 t
ORIGIN

Query Match 42.7%; Score 698.8; DB 11; Length 2573;
Best Local Similarity 65.9%; Pred. No. 4e-159;
Matches 1050; Conservative 0; Mismatches 532; Indels 12; Gaps 2;

```
Qy         19 CTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGCTCCTGCAGCTC---TTCT 75
Db         22 CTTTCAGCAAGATGTCTATGAAACAGGCTTCAGTTTTTCTGTTGATACAGTTCATATGCT 81

Qy         76 GTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTGACATGAGCCATTGGC 135
Db         82 ATATTAGACCTGGAGCCTGTGGGAAAGTGCTGGTGTGGCCTACAGAATACAGCCATTGGA 141

Qy        136 TTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATGAGGTAACAGTATTGA 195
Db        142 TAAATATGAAAATAATCCTGGATGAACTTGTCCAGAGAGGTCATGACGTCACCGTTCTCA 201

Qy        196 CTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCATTGAAATTTGAGGTGG 255
Db        202 TATCTTCTGCTTCCATCCTCATTGGGCCGAGCAATGAATCTTCTATTAATTTTGAAATTT 261

Qy        256 TCCATATGCC---------ACAGGACAGAACAGAAGAAAATGAAATATTTGTTGACCTAG 306
Db        262 ATTCTGCACCTTTGAGTAAAGATGATCTTGAATATGCTTTTGAAAAATGGGTAGGAAACT 321

Qy        307 CTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATGATTTTT 366
Db        322 GGACATACGAATTAAAAAAACTTCCATTTTGGACATCTTATTCAAAACTGCAAAAAATCT 381

Qy        367 TTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATCAGACGC 426
Db        382 CCAGTGAATATTCAGACATGATTGAAAGTTTCTGCAAAGCAGTAGTTTGGAACAAGAGCC 441

Qy        427 TTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGATTCCCT 486
Db        442 TCATGAAAAAACTCCAAGGATCTAAGTTTGATGTCGTTCTAGCAGATGCCTTGGTTCCCT 501

Qy        487 GTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAATTTCTG 546
Db        502 GTGGTGAGCTGCTATCAGAACTGCTTTAGACACCTTTTGTATACAGTCTCCGCTTCTGTC 561

Qy        547 TAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATGTACCTG 606
Db        562 CCGGATACTAATGTGAAAAGTACAGTGGGGGCCTTCCACTCCCCCCTTCCCATGTGCCTG 621

Qy        607 TGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATTCAATGC 666
Db        622 TGCTTCTGTCAGAACTAAGTGACCACATGACATTTGCAGAAAGGGTGAAGAATATGTTGC 681

Qy        667 TTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAGAGTTTT 726
Db        682 AGGTGTTGCTTTTTGACTTTTGGTTTCAAACATTTAACGAGAAATCCTGGAATCAGTTTT 741

Qy        727 ATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTGAGATAT 786
Db        742 ACAGTGATGTTCTAGGGAGACCTACAACATTAACTGAGATGATGGGGAAGGCAGACATAT 801

Qy        787 GGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACTTTGAGT 846
Db        802 GGCTCGTTCGAACCTTCTGGGACTTGAAATTTCCTCACCCTTTCTTGCCTAATTTTGACT 861

Qy        847 TTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAAATTTTG 906
Db        862 TTGTTGGAGGACTCCATTGTAAACCAGCCAAACCACTGCCTAAGGAAATGGAAGAATTTG 921

Qy        907 TCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTCAAAATG 966
Db        922 TTCAGAGCTCTGGAGAACATGGTGTGGTGGTGTTTTCTCTGGGATCAATGGTTAAAAACA 981

Qy        967 TTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGAAGGTGT 1026
Db        982 TTAAAGAAGAAAAGGCCAATGTAGTTGCTTCTGCTCTTGCCCAGATTCCACAGAAGGTTC 1041

Qy       1027 TATGGAGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGTATGATT 1086
Db       1042 TGTGGAGATTTGATGGTAAGAAACCAGACACCTTAGGATCCAACACTCGGCTCTACAAGT 1101

Qy       1087 GGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTCATGGTG 1146
Db       1102 GGATCCCCCAGAATGACCTTCTTGGTCATCCAAAAACCAAAGCTTTTATAGCTCATGGTG 1161

Qy       1147 GAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTCCCATAT 1206
Db       1162 GAACCAATGGCATCTATGAGGCGATCTACCATGGCATTCCTATTGTTGGTATTCCCTTGT 1221

Qy       1207 TTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAGAAATAA 1266
Db       1222 TTGGGGATCAACCTGATAATATTAACCACATAGTAGCCAAGGGAGCAGCTGTTAGAGTGG 1281

Qy       1267 ACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTACCGATT 1326
Db       1282 ACTTTGATACAATGTCAACTACAGACCTTCTCACTGCCTTGAAGACTGTCATTAATGACC 1341

Qy       1327 CCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTGTAAAGC 1386
Db       1342 CTTCCTATAAAGAGAACGCCATGAGATTATCCAGAATCCACCATGACCAGCCAATGAAGC 1401

Qy       1387 CCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCAAGCACC 1446
Db       1402 CCTTGGACCGAGCTGTCTTCTGGATCGAGTATGTCATGCGCAACAAGGGAGCCAAGCACC 1461

Qy       1447 TGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGATTGGGT 1506
Db       1462 TTCGCCCAGCTCTGCATGACCTTACCTGGTTCCAGTACCACTCTCTGGATGTGATTGGGT 1521

Qy       1507 TCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTATTTTCCT 1566
```

```
Db       1522 TCCTATTGGTCTGTGTGGTAGCTGTGGTATTCATCATTGCAAAATGTTGCCTCTTTTGTT 1581

Qy       1567 GTCAAAAATTTAATAAAACTAGAAAGATAGAAAA 1600
              | || ||   |  ||| |   ||||||   ||||
Db       1582 GCCATAAGACTGCTAACATGGGAAAGAAGAAAAA 1615


RESULT 3
BQ713091
LOCUS       BQ713091                960 bp    mRNA    linear   EST 16-JUL-2002
DEFINITION  AGENCOURT_8295152 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6309767
            5', mRNA sequence.
ACCESSION   BQ713091
VERSION     BQ713091.1  GI:21851990
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 960)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Susan L. Sullivan, PhD.
             cDNA Library Preparation: ResGen, Invitrogen Corp
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13730  row: p  column: 24
            High quality sequence stop: 737.
FEATURES             Location/Qualifiers
     source          1. .960
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6309767"
                     /clone_lib="NIH_MGC_129"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: olfactory epithelium; Vector:
                     pCMV-SPORT6.1.ccdb; Site_1: EcoRV; Site_2: NotI; Cloned
                     unidirectionally. Primer: Oligo dT. Average insert size
                     2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this
                     is a NIH_MGC Library."
BASE COUNT      273 a    204 c    237 g    246 t
ORIGIN

  Query Match           38.2%;  Score 625.2;  DB 14;  Length 960;
  Best Local Similarity   83.0%;  Pred. No. 3.1e-141;
  Matches  725;  Conservative    0;  Mismatches  148;  Indels .  1;  Gaps    1;

Qy        740 AGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTGAGATATGGCTAATACGAAC 799
              |||||| ||||| ||| | |||||||||| |||| |||||||||||| |||||||| |||||
Db         45 AGGAAGACCCACCACACTGTGTGAGACTATGGGGAAAGCTGAGATTTGGCTAATGCGAAC 104

Qy        800 ATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACTTTGAGTTTGTTGGAGGATT 859
              ||||||||||||||||||||||  | |||   ||| || |||||||||||| |||||| |
Db        105 CTATTGGGATTTTGAATTTCCTCGTCTATATTTACCGAATTTTGAGTTTGTGGGAGGACT 164

Qy        860 GCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAAATTTTGTCCAGAGTTCAGG 919
              |||||| ||||||||||  |||| |||||||||||||| | |||||||||  |||||
Db        165 GCACTGCAAACCTGCCAAGCCTTTACCTAAGGAAATGGAAGAGTTTGTCCAGACCTCAGG 224

Qy        920 GGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTCAAAATGTTACAGAAGAAAA 979
              |||| ||||||||||||||||||||||||||||| || |  ||||  | ||||| |||||
Db        225 GGAACATGGTATTGTGGTGTTTTCTCTGGGGTCAATGGTCAAAAACCTGACAGATGAAAA 284

Qy        980 GGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGAAGGTGTTATGGAGGTACAA 1039
              ||| ||| ||||||| |||||||||||||||| || |||||||| || ||| | |||||
Db        285 GGCCAATCTCATTGCCTCAGCCCTTGCCCAGATTCCTCAGAAGGTTTTGTGGCGATACAA 344

Qy       1040 AGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGTATGATTGGATACCCCAGAA 1099
              |||||| | ||||  |||||||||| |||||||  | |||| ||||||||| || |||||
Db        345 AGGAAAGATACCAGACACATTAGGATCCAATACAAGACTGTTTGATTGGATTCCTCAGAA 404

Qy       1100 TGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTCATGGTGGAATGAATGGGAT 1159
              |||||||||||| |||||||||||| |||||||||||| |||||||||  ||||| ||
Db        405 TGATCTTCTTGGACATCCCAAAACCAGAGCTTTTATCACCCATGGTGGAACAAATGGAAT 464

Qy       1160 CTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTCCCATATTTGGTGATCAGCT 1219
              |||||| |||||||||||||||| |||||||||||||||| ||||| || | ||| ||||
Db        465 CTATGAGGCTATTTACCATGGGATCCCTATGGTGGGAGTGCCCATGTTCGCTGACCAGCC 524

Qy       1220 TGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAGAAATAAACTTCAAAACTAT 1279
              ||| ||||| |||||||||||||||||||||||| || ||  | ||| | || || ||
Db        525 TGACAACATTGCTCACATGAAGGCCAAAGGAGCAGCAGTGGAGGTGAACATGAACACGAT 584

Qy       1280 GACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTACCGATTCCTCTTATAAAGA 1339
              ||||||| |||| | || |  ||| |||||||||||||| |  ||  | ||||||||||
Db        585 GACAAGCTCAGATCTGCTCAATGCTCTGAGAACAGTCATCAATGAGCCATCTTATAAAGA 644

Qy       1340 GAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTGTAAAGCCCCTAGATCGAGC 1399
              |||||| |||||| |||||||||| |||||||| || || || |||||||| || |||||
Db        645 GAATGCCATGAGACTATCAAGAATCCACCATGACCAGCCAGTGAAGCCCCTGGACCGAGC 704

Qy       1400 AGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCAAGCACCTGCGATCAGCTGC 1459
              |||||||||||| |||||||||||||| |||||||||||||||||||||| ||    || ||
Db        705 AGTCTTCTGGATTGAGTTTGTCATGCGTCACAAAGGAGCCAAGCACCTTCGTGTGGCAGC 764

Qy       1460 CCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGATTGGGTTCCTGCTGACCTG 1519
              |||||||||| |||||| ||| || ||||| | ||||| |||||||| ||||||  |||
Db        765 CCATGACCTCAGCTGGTTTCAGTACCACTCTCTGGATGTAATTGGGTTTCTGCTGGGCTG 824

Qy       1520 TGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTATTT-TCCTGTCAAAAATTTA 1578
              ||||||| |||||||||| ||| |  ||||||| | ||| ||| |  | ||||||| ||
Db        825 TGTGGCATCTGCTATATTATTGGTTGCAAAATGCTGTTTGTTTATAATTTCAAAAAGTTG 884

Qy       1579 ATAAAACTAGAAAGATAGAAAAGAGGGAATAGAT 1612
              ||| ||  |||||    ||||  | || ||| |
Db        885 GTAAGACAGGAAAGGAGAAAAAACGTGACTAGTT 918


RESULT 4
BQ942104
LOCUS       BQ942104                1004 bp    mRNA    linear   EST 21-AUG-2002
DEFINITION  AGENCOURT_8763113 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6313348
            5', mRNA sequence.
ACCESSION   BQ942104
VERSION     BQ942104.1  GI:22357582
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1004)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Susan L. Sullivan, PhD.
             cDNA Library Preparation: ResGen, Invitrogen Corp
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13740  row: f  column: 05
            High quality sequence stop: 656.
FEATURES             Location/Qualifiers
     source          1. .1004
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6313348"
```

/clone_lib="NIH_MGC_129"
/lab_host="DH10B (phage-resistant)"
/note="Organ: olfactory epithelium; Vector: pCMV-SPORT6.1.ccdb; Site_1: EcoRV; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this is a NIH_MGC Library."

BASE COUNT 286 a 219 c 249 g 249 t 1 others
ORIGIN

Query Match 38.1%; Score 623.2; DB 14; Length 1004;
Best Local Similarity 81.9%; Pred. No. 9.5e-141;
Matches 730; Conservative 0; Mismatches 159; Indels 2; Gaps 1;

```
Qy        714 TGGGAAGAGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGA 773
              |||||    | ||||| ||| || |||||||| ||||| ||| | |||||||||| ||||
Db         53 TGGGATTCCTATTATACTAAAGCTTTAGGAAGACCCACCACACTGTGTGAGACTATGGGG 112

Qy        774 AAAGCTGAGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAA 833
              |||||||||| |||||||| ||||| ||||||||||||||||||||  |||||   |
Db        113 AAAGCTGAGATTTGGCTAATGCGAACCTATTGGGATTTTGAATTTCCTCGTCCATATTTA 172

Qy        834 CCTAACTTTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAA 893
              || || |||||||||| |||||| |||||| |||||||||| |||| |||||||||
Db        173 CCGAATTTTGAGTTTGTGGGAGGACTGCACTGCAAACCTGCCAAGCCTTTACCTAAGGAA 232

Qy        894 ATGGAAAATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCA 953
              |||||| | |||||||||  |||||||| |||||||||||||||||||||||||||||||
Db        233 ATGGAAGAGTTTGTCCAGACCTCAGGGGAACATGGTATTGTGGTGTTTTCTCTGGGGTCA 292

Qy        954 CTGTTTCAAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATC 1013
              || |  ||||  | ||||| |||||||| ||| ||||||| ||||||||||||||||
Db        293 ATGGTCAAAAACCTGACAGATGAAAAGGCCAATCTCATTGCCTCAGCCCTTGCCCAGATT 352

Qy       1014 CCACAGAAGGTGTTATGGAGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACT 1073
              || |||||||| || ||| | |||||||||| | |||| |||||||||| |||||||
Db        353 CCTCAGAAGGTTTTGTGGCGATACAAAGGAAAGATACCAGACACATTAGGATCCAATACA 412

Qy       1074 CGGCTGTATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTT 1133
              | |||| ||||||||| || ||||||||||||||| |||||||||||| |||||||
Db        413 AGACTGTTTGATTGGATTCCTCAGAATGATCTTCTTGGACATCCCAAAACCAGAGCTTTT 472

Qy       1134 ATCACTCATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTG 1193
              ||||| ||||||||||  ||||| |||||||| |||||||||||||| ||||||||||
Db        473 ATCACCCATGGTGGAACAAATGGAATCTATGAGGCTATTTACCATGGGATCCCTATGGTG 532

Qy       1194 GGAGTTCCCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCA 1253
              ||||| ||||| || | ||| |||| ||| ||||| |||||||||||||||||||||||
Db        533 GGAGTGCCCATGTTCGCTGACCAGCCTGACAACATTGCTCACATGAAGGCCAAAGGAGCA 592

Qy       1254 GCTGTAGAAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACA 1313
              || || ||  | ||| | || || ||||||||  |||| | || |  ||| ||||||||
Db        593 GCAGTGGAGGTGAACATGAACACGATGACAAGCTCAGATCTGCTCAATGCTCTGAGAACA 652

Qy       1314 GTCATTACCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGAT 1373
              ||||| |  ||  | |||||| |||||||||| |||||| |||||||||| ||||||||
Db        653 GTCATCAATGAGCCATCTTATNAAGAGAATGCCATGAGACTATCAAGAATCCACCATGAC 712

Qy       1374 CAACCTGTAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAA 1433
              || || || |||||||| || |||||||||||||||| |||||||||||||| ||||||
Db        713 CAGCCAGTGAAGCCCCTGGACCGAGCAGTCTTCTGGATTGAGTTTGTCATGCGTCACAAA 772

Qy       1434 GGAGCCAAGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATA 1493
              |||||||||||||| ||    || |||||||||||||| |||||| ||| || |||| |
Db        773 GGAGCCAAGCACCTTCGTGTGGCAGCCCATGACCTCAGCTGGTTTCAGTACCACTCTCTG 832

Qy       1494 GATGTGATTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTT--GTTCACAAAAT 1551
                |||  |||||| | |||||| |||||||||| |||||| ||| ||  || |  |||
Db        833 AATGGAATTGGGGTTCTGCTGGCCTGTGTGGCATCTGCTAAATTATTGGGTGCCAAAAGG 892

Qy       1552 GTTTTTTATTTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGA 1602
              ||  ||  ||||  | ||||||| ||  ||| |   | || | | |||| |
Db        893 GTGGTTGGTTTTAATTTCAAAAAGTTGGTAAGAACCGGAAAAGAAAAAAAA 943
```

RESULT 5
AK002736
LOCUS AK002736 1896 bp mRNA linear HTC 19-JAN-2002
DEFINITION Mus musculus adult male kidney cDNA, RIKEN full-length enriched library, clone:0610033E06:UDP-glucuronosyltransferase 2 family, member 5, full insert sequence.
ACCESSION AK002736
VERSION AK002736.1 GI:12832933
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male kidney cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library clone:0610033E06.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I., Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R., Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T., Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H., Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G., Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M., Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H., Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N., Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C., Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D., Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P., Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L., Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S. and Hayashizaki,Y.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409 (6821), 685-690 (2001)
MEDLINE 21085660
PUBMED 11217851
REFERENCE 5 (bases 1 to 1896)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,

Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.

TITLE Direct Submission

JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Gemomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)

COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0. Second strand cDNA was prepared with the primer adapter of sequence[5' GAGAGAGAGAAGGATCCAAGAGCTCAATTAATTTAATTAAACCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.

FEATURES Location/Qualifiers

source 1. .1896
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:0610033E06"
/db_xref="MGD:MGI:1904280"
/db_xref="taxon:10090"
/clone="0610033E06"
/sex="male"
/tissue_type="kidney"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"

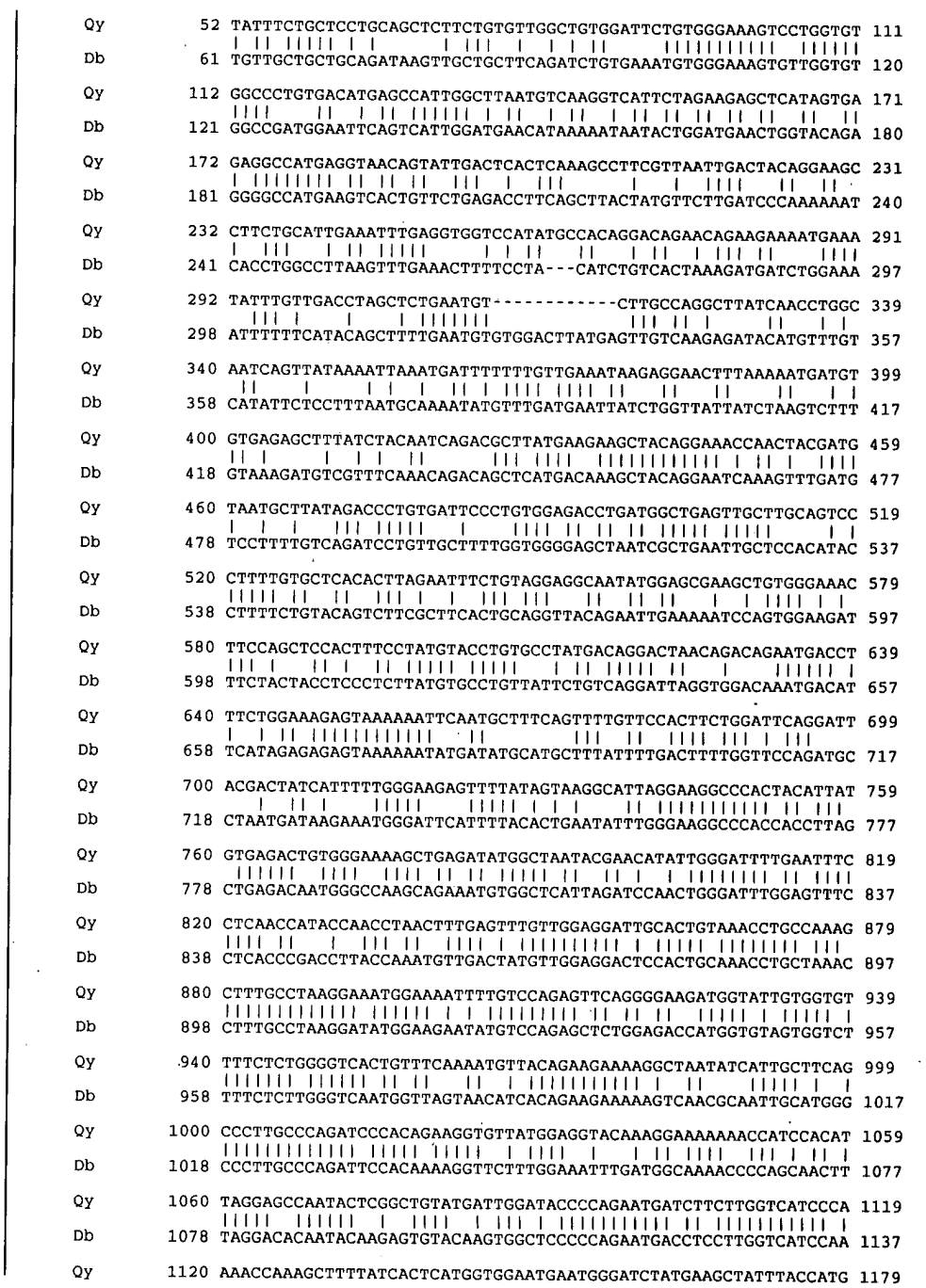
gene 1. .1896
/gene="Ugt2b5"

CDS 36. .1625
/gene="Ugt2b5"
/note="UDP-glucuronosyltransferase 2 family, member 5 data source:MGD, source key:MGI:98900, evidence:ISS putative"
/codon_start=1
/protein_id="BAB22316.1"
/db_xref="GI:12832934"
/db_xref="MGD:MGI:98900"
/translation="MPGKWMSALLLLQISCCFRSVKCGKVLVWPMEFSHWMNIKIILD ELVQRGHEVTVLRPSAYYVLDPKKSPGLKFETFPTSVTKDDLENFFIQLLNVWTYELS RDTCLSYSPLMQNMFDELSGYYLSLCKDVVSNRQLMTKLQESKFDVLLSDPVAFGGEL IAELLHIPFLYSLRFTAGYRIEKSSGRFLLPPSYVPVILSGLGGQMTFIERVKNMICM LYFDFWFQMPNDKKWDSFYTEYLGRPTTLAETMGQAEMWLIRSNWDLEFPHPTLPNVD YVGGLHCKPAKPLPKDMEEYVQSSGDHGVVVFSLGSMVSNITEEKVNAIAWALAQIPQ KVLWKFDGKTPATLGHNTRVYKWLPQNDLLGHPKTKAFITHGGANSVYEAIYHGIPMI GIPLFGEQHDNIAHMVAKGAAVTLNIRTMSRSDVLNALEEVIENPFYKKNAMWLSTIH HDQPMKPLDRAVFWVEFVMRHKNAKHLRPLGHNLTWYQYHSLDVIGFLLACVAATVVL TVKCLLFIYRFFVTKKKENEE"

BASE COUNT 571 a 361 c 386 g 578 t

ORIGIN

Query Match 38.0%; Score 622.4; DB 11; Length 1896;
Best Local Similarity 63.4%; Pred. No. 1.5e-140;
Matches 1010; Conservative 0; Mismatches 566; Indels 18; Gaps 3;



```
Db       1138 AAACCAAAGCCTTTATAACTCATGGTGGGGCCAACAGTGTCTATGAGGCAATCTATCATG 1197

Qy       1180 GGGTCCCTATGGTGGGAGTTCCCATATTTGGTGATCAGCTTGATAACATAGCTCACATGA 1239

Db       1198 GAATCCCTATGATTGGCATTCCTTTGTTTGGAGAACAGCATGATAACATTGCCCATATGG 1257

Qy       1240 AGGCCAAAGGAGCAGCTGTAGAAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGA 1299

Db       1258 TGGCCAAAGGAGCAGCTGTTACATTGAATATCAGAACAATGTCAAGGTCAGATGTGCTCA 1317

Qy       1300 GGGCTTTGAGAACAGTCATTACCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAA 1359

Db       1318 ATGCACTGGAGGAAGTCATAGAAAATCCTTTCTATAAAAAGAATGCTATGTGGTTGTCAA 1377

Qy       1360 GAATTCACCATGATCAACCTGTAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTG 1419

Db       1378 CCATTCACCATGACCAGCCTATGAAACCCCTGGACAGGGCTGTATTCTGGGTTGAGTTTG 1437

Qy       1420 TCATGCGCCACAAAGGAGCCAAGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCC 1479

Db       1438 TCATGCGCCACAAAAATGCTAAGCACTTGAGGCCACTTGGACATAACCTTACCTGGTACC 1497

Qy       1480 AGCACTACTCTATAGATGTGATTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCT 1539

Db       1498 AGTACCACTCTCTGGATGTGATTGGATTCCTACTCGCTTGTGTGGCAGCCACTGTAGTTC 1557

Qy       1540 TGTTCACAAAATGTTTTTTATTTTCCTGTCAAAAATTT---AATAAAACTAGAAAGATAG 1596

Db       1558 TTACTGTAAAGTGCCTCTTGTTCATTTACCGATTCTTTGTAACGAAAAAAAAAGAAAATG 1617

Qy       1597 AAAAGAGGGAATAGATCTTTCCAAATTCAAGAAA 1630

Db       1618 AAGAATGAGTAGAGTTCATTGTCAATGCACCACA 1651


RESULT 6
BQ925596
LOCUS       BQ925596                927 bp    mRNA    linear   EST 20-AUG-2002
DEFINITION  AGENCOURT_8888479 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6311424
            5', mRNA sequence.
ACCESSION   BQ925596
VERSION     BQ925596.1  GI:22340627
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 927)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Susan L. Sullivan, PhD.
             cDNA Library Preparation: ResGen, Invitrogen Corp
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13735  row: f  column: 01
            High quality sequence stop: 635.
FEATURES             Location/Qualifiers
     source          1. .927
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6311424"
                     /clone_lib="NIH_MGC_129"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: olfactory epithelium; Vector:
                     pCMV-SPORT6.1.ccdb; Site_1: EcoRV; Site_2: NotI; Cloned
                     unidirectionally. Primer: Oligo dT. Average insert size
                     2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this
                     is a NIH_MGC Library."
BASE COUNT      262 a    201 c    230 g    231 t      3 others
ORIGIN

  Query Match             36.7%;  Score 600;  DB 14;  Length 927;
  Best Local Similarity   80.4%;  Pred. No. 4.1e-135;
  Matches  714;  Conservative    0;  Mismatches  172;  Indels    2;  Gaps    1;

Qy        740 AGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTGAGATATGGCTAATACGAAC 799

Db         37 AGGAAGACCCACCACACTGTGTGAGACTATGGGGAAAGCTGAGATTTGGCTAATGCGAAC 96

Qy        800 ATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACTTTGAGTTTGTTGGAGGATT 859

Db         97 CTATTGGGATTTTGAATTTCCTCGTCCATATTTACCGAATTTTGAGTTTGTGGGAGGACT 156

Qy        860 GCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAAATTTTGTCCAGAGTTCAGG 919

Db        157 GCACTGCAAACCTGCCAAGCCTTTACCTAAGGAAATGGAAGAGTTTGTCCAGACCTCAGG 216

Qy        920 GGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTCAAAATGTTACAGAAGAAAA 979

Db        217 GGAACATGGTATTGTGGTGTTTTCTCTGGGGTCAATGGTCAAAAACCTGACAGATGAAAA 276

Qy        980 GGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGAAGGTGTTATGGAGGTACAA 1039

Db        277 GGCCAATCTCATTGCCTCAGCCCTTGCCCAGATTCCTCAGAAGGTTTTGTGGCGATACAA 336

Qy       1040 AGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGTATGATTGGATACCCCAGAA 1099

Db        337 AGGAAAGATACCAGACACATTAGGATCCAATACAAGACTGTTTGATTGGATTCCTCAGAA 396

Qy       1100 TGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTCATGGTGGAATGAATGGGAT 1159

Db        397 TGATCTTCTTGGACATCCCAAAACCAGAGCTTTTATCACCCATGGTGGAACAAATGGAAT 456

Qy       1160 CTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTCCCATATTTGGTGATCAGCT 1219

Db        457 CTATGAGGCTATTTACCATGGGATCCCTATGGTGGGAGTGCCCATGTTCGCTGACCAGCC 516

Qy       1220 TGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAGAAATAAACTTCAAAACTAT 1279

Db        517 TGACAACATTGCTCACATGAAGGCCAAAGGAGCAGCAGTGGAGGTGAACATGAACACGAT 576

Qy       1280 GACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTACCGATTCCTCTTATAAAGA 1339

Db        577 GACAAGCTCAGATCTGCTCAATGCTCTGAGAACAGTCATCAATGAGCCATCTTATNAAGA 636

Qy       1340 GAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTGTAAAGCCCCTAGATCGAGC 1399

Db        637 GAATGCCATGAGACTATCAAGAATCCACCATGACCAGCCAGTGAAGCCCCTGGACCGAGC 696

Qy       1400 AGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCAAGCACCTGCGATCAGCTGC 1459

Db        697 AGTCTTCTGGATTGAGTTTGTCATGCGTCACANAGGAGCCCAGCACCTTCGTGTGGCAGC 756

Qy       1460 CCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGATTGGGTTCCTGCTGACCTG 1519

Db        757 CCATGACCTCAGCTGGTTTCAGTACCACTCTCCGGATGGAATTGGGTTTCTGCTGGCCTG 816

Qy       1520 TGTGGCAACTGCTATATTCTT--GTTCACAAAATGTTTTTTATTTTCCTGTCAAAAATTT 1577

Db        817 TGTGGCATCTGCTATAATATTGGGTTGCCAAAGGGTGGTTTGCGTTAAAATTTCCAAAAA 876

Qy       1578 AATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAAATTCA 1625

Db        877 AGTTGGGTTAGGAACCGGGGAAAGAAAGAAAAAAACCGTGGGACTTAA 924


RESULT 7
BQ934372
LOCUS       BQ934372                930 bp    mRNA    linear   EST 21-AUG-2002
```

```
DEFINITION  AGENCOURT_8764945 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6314635
            5', mRNA sequence.
ACCESSION   BQ934372
VERSION     BQ934372.1  GI:22349755
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 930)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Susan L. Sullivan, PhD.
             cDNA Library Preparation: ResGen, Invitrogen Corp
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13743  row: k  column: 20
            High quality sequence stop: 547.
FEATURES             Location/Qualifiers
     source          1. .930
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6314635"
                     /clone_lib="NIH_MGC_129"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: olfactory epithelium; Vector:
                     pCMV-SPORT6.1.ccdb; Site_1: EcoRV; Site_2: NotI; Cloned
                     unidirectionally. Primer: Oligo dT. Average insert size
                     2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this
                     is a NIH_MGC Library."
BASE COUNT      275 a    200 c    230 g    225 t
ORIGIN

  Query Match           33.8%;  Score 552.4;  DB 14;  Length 930;
  Best Local Similarity   81.5%;  Pred. No. 1.5e-123;
  Matches  665;  Conservative    0;  Mismatches  146;  Indels    5;  Gaps    2;

Qy        717 GAAGAGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAA 776
              | | | ||   |  |||      |||||| ||||| ||| | ||||||||| |||| |||
Db          3 GGACACTTGGGTTTTAAAAGTACAGGAAGACCCACCACACTGTGTGAGACTATGGGGAAA 62

Qy        777 GCTGAGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCT 836
              |||||||| |||||||| ||||| ||||||||||||||||||||||  |||||   |||
Db         63 GCTGAGATTTGGCTAATGCGAACCTATTGGGATTTTGAATTTCCTCGTCCATATTTACCG 122

Qy        837 AACTTTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATG 896
              || |||||||||| |||||| ||||||| ||||||||||  |||| ||||||||||||
Db        123 AATTTTGAGTTTGTGGGAGGACTGCACTGCAAACCTGCCAAGCCTTTACCTAAGGAAATG 182

Qy        897 GAAAATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTG 956
              ||| | ||||||||||  |||||||| ||||||||||||||||||||||||||||||||
Db        183 GAAGAGTTTGTCCAGACCTCAGGGGAACATGGTATTGTGGTGTTTTCTCTGGGGTCAATG 242

Qy        957 TTTCAAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCA 1016
                |  ||||  | ||||| |||||||| ||| ||||||| ||||||||||||||||| ||
Db        243 GTCAAAAACCTGACAGATGAAAAGGCCAATCTCATTGCCTCAGCCCTTGCCCAGATTCCT 302

Qy       1017 CAGAAGGTGTTATGGAGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGG 1076
              |||||||| || ||| | |||||||||||| | ||||  |||||||||| |||||||  |
Db        303 CAGAAGGTTTTGTGGCGATACAAAGGAAAGATACCAGACACATTAGGATCCAATACAAGA 362

Qy       1077 CTGTATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATC 1136
              |||| ||||||||| || |||||||||||||||| |||||||||||| ||||||||||
Db        363 CTGTTTGATTGGATTCCTCAGAATGATCTTCTTGGACATCCCAAAACCAGAGCTTTTATC 422

Qy       1137 ACTCATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGA 1196
              || |||||||||  ||||| |||||||| |||||||||||||| ||||||||||||||
Db        423 ACCCATGGTGGAACAAATGGAATCTATGAGGCTATTTACCATGGGATCCCTATGGTGGGA 482

Qy       1197 GTTCCCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCT 1256
              || ||||| || | ||| |||| ||| ||||| ||||||||||||||||||||||||||
Db        483 GTGCCCATGTTCGCTGACCAGCCTGACAACATTGCTCACATGAAGGCCAAAGGAGCAGCA 542

Qy       1257 GTAGAAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTC 1316
              || ||  | ||| | || || |||||||||  |||| | || |  ||| ||||||||||
Db        543 GTGGAGGTGAACATGAACACGATGACAAGCTCAGATCTGCTCAATGCTCTGAGAACAGTC 602

Qy       1317 ATTACCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAA 1376
              || |  ||  | |||||||||||||||| |||||| |||||||||| |||||||| ||
Db        603 ATCAATGAGCCATCTTATAAAGAGAATGCCATGAGACTATCAAGAATCCACCATGACCAG 662

Qy       1377 CCTGTAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGA 1436
              || || |||||||| || |||||||||||||||| |||||||||||||| |||||||||
Db        663 CCAGTGAAGCCCCTGGACCGAGCAGTCTTCTGGATTGAGTTTGTCATGCGTCACAAAGGA 722

Qy       1437 GCCAAGCACCTGCGATCAGCTGCCCAT-GACCTCACCTGGTTCC----AGCACTACTCTA 1491
              |||||||||| ||    |  ||||| | ||||| |||||| |    | | || ||||
Db        723 GCCAAGCACCTTCGTGTGGGCACCCATGGCCCTCAGCTGGTTTCCAGAACCCCTCCTCTG 782

Qy       1492 TAGATGTGATTGGGTTCCTGCTGACCTGTGTGGCAA 1527
                |  |       |||||||||||| |||||| ||| |
Db        783 GATGTAAATGGGGGTTCCTGCTGGCCTGTGGGGCCA 818


RESULT 8
BQ714919
LOCUS       BQ714919                1030 bp    mRNA    linear   EST 16-JUL-2002
DEFINITION  AGENCOURT_8480154 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6310837
            5', mRNA sequence.
ACCESSION   BQ714919
VERSION     BQ714919.1  GI:21853818
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1030)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Susan L. Sullivan, PhD.
             cDNA Library Preparation: ResGen, Invitrogen Corp
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13733  row: m  column: 14
            High quality sequence stop: 599.
FEATURES             Location/Qualifiers
     source          1. .1030
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6310837"
                     /clone_lib="NIH_MGC_129"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: olfactory epithelium; Vector:
                     pCMV-SPORT6.1.ccdb; Site_1: EcoRV; Site_2: NotI; Cloned
                     unidirectionally. Primer: Oligo dT. Average insert size
                     2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this
                     is a NIH_MGC Library."
BASE COUNT      295 a    229 c    257 g    248 t      1 others
ORIGIN

  Query Match           31.8%;  Score 520.4;  DB 14;  Length 1030;
  Best Local Similarity   78.5%;  Pred. No. 8.8e-116;
  Matches  724;  Conservative    0;  Mismatches  186;  Indels   12;  Gaps    8;
```

```
Qy        717 GAAGAGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAA 776
              | | | ||   |   || | |   ||||| ||||| ||| | |||||||||| |||| |||
Db         26 GGACACTTGGTTTTAAAAGTACAGGGAAGACCCACCACACTGTGTGAGACTATGGGGAAA 85

Qy        777 GCTGAGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCT 836
              |||||||| |||||||| ||||| |||||||||||||||||||||||  |||||   |||
Db         86 GCTGAGATTTGGCTAATGCGAACCTATTGGGATTTTGAATTTCCTCGTCCATATTTACCG 145

Qy        837 AACTTTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATG 896
              || |||||||||| |||||| ||||||| |||||||||  |||| ||||||||||||
Db        146 AATTTTGAGTTTGTGGGAGGACTGCACTGCAAACCTGCCAAGCCTTTACCTAAGGAAATG 205

Qy        897 GAAAATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTG 956
              ||| | ||||||||||  |||||||| ||||||||||||||||||||||||||||||| ||
Db        206 GAAGAGTTTGTCCAGACCTCAGGGGAACATGGTATTGTGGTGTTTTCTCTGGGGTCAATG 265

Qy        957 TTTCAAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCA 1016
                |  ||||  | ||||| |||||||| ||| ||||||| |||||||||||||||| ||
Db        266 GTCAAAAACCTGACAGATGAAAAGGCCAATCTCATTGCCTCAGCCCTTGCCCAGATTCCT 325

Qy       1017 CAGAAGGTGTTATGGAGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGG 1076
              ||||||| || ||| | |||||||||| | ||||  |||||||||  |||||||  |
Db        326 CAGAAGGTTTTGTGGCGATACAAAGGAAAGATACCAGACACATTAGGATCCAATACAAGA 385

Qy       1077 CTGTATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATC 1136
              |||| |||||||| || ||||||||||||||| |||||||||||| ||||||||||
Db        386 CTGTTTGATTGGATTCCTCAGAATGATCTTCTTGGACATCCCAAAACCAGAGCTTTTATC 445

Qy       1137 ACTCATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGA 1196
              || ||||||||||  ||||| |||||||| |||||||||||||| ||||||||||||||
Db        446 ACCCATGGTGGAACAAATGGAATCTATGAGGCTATTTACCATGGGATCCCTATGGTGGGA 505

Qy       1197 GTTCCCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCT 1256
              || ||||| || | ||| |||| ||| ||||| ||||||||||||||||||||||||||
Db        506 GTGCCCATGTTCGCTGACCAGCCTGACAACATTGCTCACATGAAGGCCAAAGGAGCAGCA 565

Qy       1257 GTAGAAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTC 1316
              || ||  | ||| | || || |||||||||  |||| | || |  ||| ||||||||||
Db        566 GTGGAGGTGAACATGAACACGATGACAAGCTCAGATCTGCTCAATGCTCTGAGAACAGTC 625

Qy       1317 ATTACCGATTCC-TCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCA 1375
              || |  ||  || |||||||||||||||| |||||| |||||||||| | |||||| ||
Db        626 ATCAATGAGCCCATCTTATAAAGAGAATGCCATGAGACTATCAAGAATCCCCCATGACCA 685

Qy       1376 ACCTGTAAAGCCCCTAGA-TCGAGCAGTCTTCTGGATCGAG-TTTGTCATGCGCCACAAA 1433
              || || |||||||| ||  |||||||||||||||| ||| |||||||||| ||||||
Db        686 GCCAGTGAAGCCCCTGGACCCGAGCAGTCTTCTGGATTGAGTTTTGTCATGCGTCACAAA 745

Qy       1434 GG-AGCCAAGCACCTGCGATCAGCTGCCCATGACCTCACCT--GGTTCCAGCACTACTCT 1490
              || |||||||||||| ||    || ||||||||||    ||  ||||| ||  | |||||
Db        746 GGAAGCCAAGCACCTTCGGGTGGCAGCCCATGACCCTCACTTGGGTTCAAGACCCACTCT 805

Qy       1491 ATAGATGTGA--TTGGGTTCCTGCTGACCTGTGTGGCAA-CTGCTATATTCTTGTT---C 1544
                | ||||  |  | || |||   |||  |||||||||||| |||||||||| ||| |   |
Db        806 CTGGATGGAAATTGGGTTTCTGCCTGGGCTGTGTGGCAATCTGCTATATTATTGGTTGGC 865

Qy       1545 ACAAAATGTTTTTTATTTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGG 1604
              | ||      ||   || |  | ||  ||| ||  ||| ||  ||||||   |||| |
Db        866 AAAAGGGTGGTTGGTTTATATTCTCCCAAAGTTGGTAACACCGGAAAGAAGAAAAAAACT 925

Qy       1605 GAATAGATCTTTCCAAATTCAA 1626
              | |   || |    |||   ||
Db        926 GGACTAATTTGAAGAAAAAAAA 947


RESULT 9
BQ714368
LOCUS       BQ714368                913 bp    mRNA    linear   EST 16-JUL-2002
DEFINITION  AGENCOURT_8468705 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6308241
            5', mRNA sequence.
ACCESSION   BQ714368
VERSION     BQ714368.1  GI:21853267
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 913)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Susan L. Sullivan, PhD.
             cDNA Library Preparation: ResGen, Invitrogen Corp
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13727  row: a  column: 10
            High quality sequence stop: 654.
FEATURES             Location/Qualifiers
     source          1. .913
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6308241"
                     /clone_lib="NIH_MGC_129"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: olfactory epithelium; Vector:
                     pCMV-SPORT6.1.ccdb; Site_1: EcoRV; Site_2: NotI; Cloned
                     unidirectional. Primer: Oligo dT. Average insert size
                     2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this
                     is a NIH_MGC Library."
BASE COUNT      258 a    221 c    203 g    231 t
ORIGIN

  Query Match           30.1%;  Score 492.6;  DB 14;  Length 913;
  Best Local Similarity   81.0%;  Pred. No. 5e-109;
  Matches  573;  Conservative    0;  Mismatches  134;  Indels    0;  Gaps    0;

Qy        919 GGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTCAAAATGTTACAGAAGAAA 978
              || |   | |   |||||||||||||||||||||  || |  |||  | ||||| ||||
Db          4 GGAATTCTCGGGATGTGGTGTTTTCTCTGGGGTCTATGGTCAAAAACCTGACAGATGAAA 63

Qy        979 AGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGAAGGTGTTATGGAGGTACA 1038
              |||| ||| ||||||| |||||||||||||||| || |||||||| || ||| | ||||
Db         64 AGGCCAATCTCATTGCCTCAGCCCTTGCCCAGATTCCTCAGAAGGTTTTGTGGCGATACA 123

Qy       1039 AAGGAAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGTATGATTGGATACCCCAGA 1098
              ||||||| | ||||  |||||||||| |||||||  | |||| ||||||||| || ||||
Db        124 AAGGAAAGATACCAGACACATTAGGATCCAATACAAGACTGTTTGATTGGATTCCTCAGA 183

Qy       1099 ATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTCATGGTGGAATGAATGGGA 1158
              |||||||||||| |||||||||||| |||||||||| |||||||||  ||||| |
Db        184 ATGATCTTCTTGGACATCCCAAAACCAGAGCTTTTATCACCCATGGTGGAACAAATGGAA 243

Qy       1159 TCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTCCCATATTTGGTGATCAGC 1218
              ||||||| |||||||||||||| |||||||||||||||| ||||| || | ||| ||||
Db        244 TCTATGAGGCTATTTACCATGGGATCCCTATGGTGGGAGTGCCCATGTTCGCTGACCAGC 303

Qy       1219 TTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAGAAATAAACTTCAAAACTA 1278
              ||| ||||| |||||||||||||||||||||||||| || ||  | ||| | || || |
Db        304 CTGACAACATTGCTCACATGAAGGCCAAAGGAGCAGCAGTGGAGGTGAACATGAACACGA 363

Qy       1279 TGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTACCGATTCCTCTTATAAAG 1338
              |||||||  |||| | || |  ||| ||||||||||||| |  ||  | ||||||||||
Db        364 TGACAAGCTCAGATCTGCTCAATGCTCTGAGAACAGTCATCAATGAGCCATCTTATAAAG 423

Qy       1339 AGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTGTAAAGCCCCTAGATCGAG 1398
              ||||||| |||||| |||||||||| |||||||| || || || |||||||| || ||||
Db        424 AGAATGCCATGAGACTATCAAGAATCCACCATGACCAGCCAGTGAAGCCCCTGGACCGAG 483

Qy       1399 CAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCAAGCACCTGCGATCAGCTG 1458
              |||||||||||||| |||||||||||||| |||||||||||||||||||||  ||    || |
Db        484 CAGTCTTCTGGATTGAGTTTGTCATGCGTCACAAAGGAGCCAAGCACCTTCGTGTGGCAG 543
```

```
Qy       1459 CCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGATTGGGTTCCTGCTGACCT 1518
             |||||||||||| |||||| ||| || ||||| | ||||| |||||||| |||||| |||
Db        544 CCCATGACCTCAGCTGGTTTCAGTACCACTCTCTGGATGTAATTGGGTTTCTGCTGGCCT 603

Qy       1519 GTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTATTTTCCTGTCAAAAATTTA 1578
             |||||||| |||||||||| ||| |  |||||||| ||| |||   | |||||||  ||
Db        604 GTGTGGCATCTGCTATATTATTGGTTGCAAAATGTTGTTTGTTTATATTTCAAAAAGTTG 663

Qy       1579 ATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAAATTCA 1625
              ||| ||  ||||||   ||||  | || ||| |     |||| ||
Db        664 GTAAGACAGGAAAGAAGAAAAAACGTGACTAGTTGAAGACAAAGACA 710


RESULT 10
BF689099/c
LOCUS       BF689099                 823 bp    mRNA    linear   EST 22-DEC-2000
DEFINITION  602185172T1 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:4299880 3',
            mRNA sequence.
ACCESSION   BF689099
VERSION     BF689099.1  GI:11974507
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 823)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1157  row: o  column: 17
            High quality sequence start: 5
            High quality sequence stop: 796.
FEATURES             Location/Qualifiers
     source          1. .823
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4299880"
                     /clone_lib="NIH_MGC_43"
                     /tissue_type="normal pigmented retinal epithelium"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and  Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library. |"
BASE COUNT      221 a    199 c    164 g    239 t
ORIGIN

  Query Match           29.2%;  Score 477;  DB 12;  Length 823;
  Best Local Similarity   78.0%;  Pred. No. 3.1e-105;
  Matches  587;  Conservative    0;  Mismatches  165;  Indels    1;  Gaps    1;

Qy        865 GTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAAATTTTGTCCAGAGTTCAGGGGAAG 924
              | ||||| |||||  | |  |  | |||||||||| | ||||| |||||||| || |||
Db        804 GCAAACCGGCCAACCCCTGCCTAAGGGAAATGGAAGACTTTGTACAGAGTTCTGGAGAAA 745

Qy        925 ATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTCAAAATGTTACAGAAGAAAAGGCTA 984
              ||||| |||||||||||||||||||||| || |    ||  | |||||||||| ||| |
Db        744 ATGGTGTTGTGGTGTTTTCTCTGGGGTCAATGGTCAGTAACATGACAGAAGAAAGGGCCA 685

Qy        985 ATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGAAGGTGTTATGGAGGTACAAAGGAA 1044
              |  | ||||| |||| ||| |||||||||||||| |||||  | |||||  |  || |
Db        684 ACGTAATTGCATCAGACCTGGCCCAGATCCCACAAAAGGTTCTGTGGAGATTTGATGGGA 625

Qy       1045 AAAAACCATCCACATTAGGAGCCAATACTCGGCTGTATGATTGGATACCCCAGAATGATC 1104
              | ||||||   || |||||  |||||||||||||||| | |||||||||||||||| |
Db        624 ATAAACCAGATACCTTAGGTCTCAATACTCGGCTGTATAAGTGGATACCCCAGAATGACC 565

Qy       1105 TTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTCATGGTGGAATGAATGGGATCTATG 1164
              |||| |||||||| || |||| |||||||| |||||||||||   ||||| ||||| |
Db        564 TTCTAGGTCATCCAAAGACCAGAGCTTTTATAACTCATGGTGGAGCCAATGGCATCTACG 505

Qy       1165 AAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTCCCATATTTGGTGATCAGCTTGATA 1224
              | || || |||||||| |||||||||||||  ||||  | ||||  ||||| | |||||
Db        504 AGGCAATCTACCATGGGATCCCTATGGTGGGGATTCCATTGTTTGCCGATCAACCTGATA 445

Qy       1225 ACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAGAAATAAACTTCAAAACTATGACAA 1284
              |||| ||||||||||||||||  ||||||||||   | |  |||||| || ||| | |
Db        444 ACATTGCTCACATGAAGGCCAGGGGAGCAGCTGTTAGAGTGGACTTCAACACAATGTCGA 385

Qy       1285 GCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTACCGATTCCTCTTATAAAGAGAATG 1344
              |   ||| || ||||  || ||||  | ||| ||||  ||| | || ||||||||||||
Db        384 GTACAGACTTGCTGAATGCATTGAAGAGAGTAATTAATGATCCTTCATATAAAGAGAATG 325

Qy       1345 CTATGAGATTATCAAGAATTCACCATGATCAACCTGTAAAGCCCCTAGATCGAGCAGTCT 1404
              ||||| |||||||||||||| |||||||||| || ||| |||| ||||||||||||
Db        324 TTATGAAATTATCAAGAATTCAACATGATCAACCAGTGAAG-CCCTGGATCGAGCAGTCT 266

Qy       1405 TCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCAAGCACCTGCGATCAGCTGCCCATG 1464
              ||||||| || |||||||||||||||||||||||| || ||||| ||    || ||||| |
Db        265 TCTGGATTGAATTTGTCATGCGCCACAAAGGAGCTAAACACCTTCGGGTTGCAGCCCACG 206

Qy       1465 ACCTCACCTGGTTCCAGCACTACTCTATAGATGTGATTGGGTTCCTGCTGACCTGTGTGG 1524
              |||||||||||||||| || ||||| | |||||||||||||||||||  ||||||||
Db        205 ACCTCACCTGGTTCCAGTACCACTCTTTGGATGTGATTGGGTTCCTGCTGGTCTGTGTGG 146

Qy       1525 CAACTGCTATATTCTTGTTCACAAAATGTTTTTTATTTTCCTGTCAAAAATTTAATAAAA 1584
              ||||||  |||||  | |||||||||||| | | ||||  |      || |||  || ||
Db        145 CAACTGTGATATTTATCGTCACAAAATGTTGTCTGTTTTGTTTCTGGAAGTTTGCTAGAA 86

Qy       1585 CTAGAAAGATAGAAAAGAGGGAATAGATCTTTC 1617
              |||||  | ||| |  || ||| | | ||
Db         85 AAGCAAAGAAGGGAAAAAATGATTAGTTATATC 53


RESULT 11
BG249053
LOCUS       BG249053                 671 bp    mRNA    linear   EST 13-FEB-2001
DEFINITION  602361828F1 NIH_MGC_89 Homo sapiens cDNA clone IMAGE:4470199 5',
            mRNA sequence.
ACCESSION   BG249053
VERSION     BG249053.1  GI:12758869
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 671)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10287  row: h  column: 08
            High quality sequence stop: 661.
FEATURES             Location/Qualifiers
```

```
     source          1. .671
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4470199"
                     /clone_lib="NIH_MGC_89"
                     /tissue_type="hypernephroma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      206 a    123 c    153 g    189 t
ORIGIN

  Query Match           28.6%;  Score 467.4;  DB 12;  Length 671;
  Best Local Similarity   94.2%;  Pred. No. 6.6e-103;
  Matches  606;  Conservative    0;  Mismatches   21;  Indels   16;  Gaps   11;

Qy        193 TGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCATTGAAATTTGAGG 252
              ||||||||||||||||||||||| ||||||||||||||||||||||  ||  ||||||
Db         11 TGACTCACTCAAAGCCTTCGTTAA-TGACTACAGGAAGCCTTCTGCATGAAA--TTGAGG 67

Qy        253 TGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTGACCTAGCTCTGA 312
              |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db         68 TGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATA-TTGTTGACCTAGCTCTGA 126

Qy        313 ATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATGATTTTTTTGTTG 372
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        127 ATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATGATTTTTTTGTTG 186

Qy        373 AAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATCAGACGCTTATGA 432
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        187 AAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATCAGACGCTTATGA 246

Qy        433 AGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGATTCCCTGTGGAG 492
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        247 AGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGATTCCCTGTGGAG 306

Qy        493 ACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAATTTCT-GTAGGA 551
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        307 ACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAATTTCTGGTAGGA 366

Qy        552 GGCAATATGGAGCGAAGCT-GTGGGAAACTTCCAGCTCCA-CTTTCCTATGTACCTGTGC 609
              |||||||||||||||||| |||||||||||||||||||| ||||||||||||||||||
Db        367 GGCAATATGGAGCGAAGCTGGTGGGAAACTTCCAGCTCCAGCTTTCCTATGTACCTGTGC 426

Qy        610 CTATGACAGGACTAACAGACAGAA-TGACCTTTCTGGAAAGAGT---AAAAAATTCAATG 665
              |||||||||||||||||||||||| ||||||||||||||||||   | || ||||||||
Db        427 CTATGACAGGACTAACAGACAGAAGTGACCTTTCTGGAAAGAGTCAAACAACATTCAATG 486

Qy        666 CTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAGAGTTT 725
              |||||||||||| |||||||||||||||||||||||||||| ||||| |    |
Db        487 CTTTCAGTTTTGTGCCACTTCTGGATTCAGGATTACGACTATCATTGTTGGGGAAGAGTG 546

Qy        726 T--ATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTGAGA 783
              |   |||||||||| |  | | | ||||||||||||||||||||||||||||||||||
Db        547 TGAGTAGTAAGGCAATTAGGAAGGCCACTACATTATGTGAGACTGTGGGAAAAGCTGAGA 606

Qy        784 TATGG-CTAATACGAACATATTGGGATTTTGAATTTCCTCAAC 825
              |||||  ||||||||||||||  ||  ||||||||||||||||
Db        607 TATGGCCTAATACGAACATAT--GGGATTTGAATTTCCTCAAC 647


RESULT 12
BI331895
LOCUS       BI331895                977 bp    mRNA    linear   EST 30-JUL-2001
DEFINITION  602982580F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:5135491 5',
            mRNA sequence.
ACCESSION   BI331895
VERSION     BI331895.1  GI:15016552
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 977)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11332  row: p  column: 20
            High quality sequence stop: 873.
FEATURES             Location/Qualifiers
     source          1. .977
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5135491"
                     /clone_lib="NCI_CGAP_Li9"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.9 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      267 a    218 c    222 g    270 t
ORIGIN

  Query Match           28.4%;  Score 464.4;  DB 13;  Length 977;
  Best Local Similarity   72.5%;  Pred. No. 3.5e-102;
  Matches  657;  Conservative    0;  Mismatches  241;  Indels    8;  Gaps    4;

Qy        633 ATGACCTTTCTGGAAAGAGTAAAAAATTCAATGCTTTCAGTTTTGTTCCACTTCTGGATT 692
              ||||| |||   ||||| || || |||    |||      |  | ||  |||| ||| ||
Db          1 ATGACATTTGCAGAAAGGGTGAAGAATATGTTGCAGGTGTTGCTTTTTGACTTTTGGTTT 60

Qy        693 CAGGATTACGACTATCATTTTTGGGAAGAGTTTTATAGTAAGGCATTAGGAAGGCCCACT 752
              ||    |  || |  | |  ||| |  |||||| ||| | |   |||| || || ||
Db         61 CAAACATTTAACGAGAAATCCTGGAATCAGTTTTACAGTGATGTTCTAGGGAGACCTACA 120

Qy        753 ACATTATGTGAGACTGTGGGAAAAGCTGAGATATGGCTAATACGAACATATTGGGATTTT 812
              ||||||  |||||   |||| || || || ||||||||  | |||||  |  ||||| ||
Db        121 ACATTAACTGAGATGATGGGGAAGGCAGACATATGGCTCGTTCGAACCTTCTGGGACTTG 180

Qy        813 GAATTTCCTCAACCATACCAACCTAACTTTGAGTTTGTTGGAGGATTGCACTGTAAACCT 872
              |||||||||| || | |  ||||| ||||| |||||||||||| | || ||||||||
Db        181 AAATTTCCTCACCCTTTCTTGCCTAATTTTGACTTTGTTGGAGGACTCCATTGTAAACCA 240

Qy        873 GCCAAAGCTTTGCCTAAGGAAATGGAAAATTTTGTCCAGAGTTCAGGGGAAGATGGTATT 932
              ||||||  |||||||||||||||| | ||||| ||||| || || ||| ||||| |
Db        241 GCCAAACCACTGCCTAAGGAAATGGAAGAATTTGTTCAGAGCTCTGGAGAACATGGTGTG 300

Qy        933 GTGGTGTTTTCTCTGGGGTCACTGTTTCAAAATGTTACAGAAGAAAAGGCTAATATCATT 992
              |||||||||||||||| ||| || || |||| ||| ||||||||||| ||| |  ||
Db        301 GTGGTGTTTTCTCTGGGATCAATGGTTAAAAACATTAAAGAAGAAAAGGCCAATGTAGTT 360

Qy        993 GCTTCAGCCCTTGCCCAGATCCCACAGAAGGTGTTATGGAGGTACAAAGGAAAAAAACCA 1052
              ||||| || |||||||||| ||||||||||  | ||||| |   | || || ||||||
Db        361 GCTTCTGCTCTTGCCCAGATTCCACAGAAGGTTCTGTGGAGATTTGATGGTAAGAAACCA 420

Qy       1053 TCCACATTAGGAGCCAATACTCGGCTGTATGATTGGATACCCCAGAATGATCTTCTTGGT 1112
                ||| |||||| ||| |||||||| ||  | ||||| |||||||||| |||||||||
Db        421 GACACCTTAGGATCCAACACTCGGCTCTACAAGTGGATCCCCCAGAATGACCTTCTTGGT 480

Qy       1113 CATCCCAAAACCAAAGCTTTTATCACTCATGGTGGAATGAATGGGATCTATGAAGCTATT 1172
              ||||| ||||||||||||||||  |||||||||||  ||||| |||||||| || ||
Db        481 CATCCAAAAACCAAAGCTTTTATAGCTCATGGTGGAACCAATGGCATCTATGAGGCGATC 540
```

```
Qy       1173 TACCATGGGGTCCCTATGGTGGGAGTTCCCATATTTGGTGATCAGCTTGATAACATAGCT 1232
             ||||||||  | ||||| || ||  ||||| | ||||| ||||| | |||||| ||
Db        541 TACCATGGCATTCCTATTGTTGGTATTCCCTTGTTTGGGGATCAACCTGATAATATTAAC 600

Qy       1233 CACATGAAGGCCAAAGGAGCAGCTGTAGAAATAAACTTCAAAACTATGACAAGCGAAGAT 1292
             |||||    ||||| ||||||||||||   | |  ||||  | || ||| |||    |||
Db        601 CACATAGTAGCCAAGGGAGCAGCTGTTAGAGTGGACTTTGATACAATGTCAACTACAGAC 660

Qy       1293 TTACTGAGGGCTTTGAGAACAGTCATTACCGATTCCTCTTATAAAGAGAATGCTATGAGA 1352
               | || |  || ||||  ||.||||||  ||  | || ||||||||||||||| ||||||
Db        661 CTTCTCACTGCCTTGAAGACTGTCATTAATGACCCTTCCTATAAAGAGAATGCCATGAGA 720

Qy       1353 TTATCAAGAATTCACCATGATCA-ACCTGTAAAGCCCCTAGATCGAGCAGTCTTCTGGAT 1411
             |||||  |||||  |||||||| || |||   | |||||| | || |||||  |||||||||
Db        721 TTATCCAGAATCCACCATGACCAGACCAATGAAGCCCTTGGACCGAGCTGTCTTCTGGGA 780

Qy       1412 CGAGTTTGT---CATGCGCCACAAAGGAGCCAA---GCACCTGCGATCAGCTGCCCATG- 1464
                 | |   ||||||| |||| ||| ||||       | | ||  |||||   ||||
Db        781 TCGACGTATGTCCATGCGCAACAAGGGAACCAAAGCAACCTTTCGCCCAGCTCTGCATGA 840

Qy       1465 ACCTCACCTGGTTCCAGCACTACTCTATAGATGTGATTGGGTTCCTGCTGACCTGTGTGG 1524
             ||||  |||||||||||| || ||||| | ||||      || ||||   | |||| ||||
Db        841 ACCTTACCTGGTTCCAGTACCACTCTCTGGATGGTGACTGGGTCCTATAGGCCTGGGTGG 900

Qy       1525 CAACTG 1530
                 | |||
Db        901 TACCTG 906
```

RESULT 13
BQ900474
LOCUS BQ900474 871 bp mRNA linear EST 16-AUG-2002
DEFINITION AGENCOURT_8765257 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6312552 5', mRNA sequence.
ACCESSION BQ900474
VERSION BQ900474.1 GI:22292495
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 871)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Susan L. Sullivan, PhD.
cDNA Library Preparation: ResGen, Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLAM13738 row: e column: 01
High quality sequence stop: 671.
FEATURES Location/Qualifiers
source 1. .871
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:6312552"
/clone_lib="NIH_MGC_129"
/lab_host="DH10B (phage-resistant)"
/note="Organ: olfactory epithelium; Vector: pCMV-SPORT6.1.ccdb; Site_1: EcoRV; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this is a NIH_MGC Library."
BASE COUNT 251 a 202 c 191 g 227 t
ORIGIN

Query Match 28.4%; Score 464; DB 14; Length 871;
Best Local Similarity 81.2%; Pred. No. 4.4e-102;

Matches 539; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

```
Qy        972 GAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGAAGGTGTTATGG 1031
              || |||||||| ||| ||||||| ||||||||||||||||||| || |||||||| || |||
Db         17 GATGAAAAGGCCAATCTCATTGCCTCAGCCCTTGCCCAGATTCCTCAGAAGGTTTTGTGG 76

Qy       1032 AGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGTATGATTGGATA 1091
               | |||||||||||| | ||||  |||||||||| |||||||  | |||| |||||||||
Db         77 CGATACAAAGGAAAGATACCAGACACATTAGGATCCAATACAAGACTGTTTGATTGGATT 136

Qy       1092 CCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTCATGGTGGAATG 1151
              || |||||||||||||||| |||||||||||| |||||||||||| ||||||||||
Db        137 CCTCAGAATGATCTTCTTGGACATCCCAAAACCAGAGCTTTTATCACCCATGGTGGAACA 196

Qy       1152 AATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTCCCATATTTGGT 1211
              ||||| |||||||| |||||||||||||||| |||||||||||||||||| ||||| || | |
Db        197 AATGGAATCTATGAGGCTATTTACCATGGGATCCCTATGGTGGGAGTGCCCATGTTCGCT 256

Qy       1212 GATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAGAAATAAACTTC 1271
              || |||| ||| ||||| |||||||||||||||||||||||||| || ||  | ||| |
Db        257 GACCAGCCTGACAACATTGCTCACATGAAGGCCAAAGGAGCAGCAGTGGAGGTGAACATG 316

Qy       1272 AAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTACCGATTCCTCT 1331
              || || |||||||||  |||| | || |  ||| ||||||||||||| |  ||  | |||
Db        317 AACACGATGACAAGCTCAGATCTGCTCAATGCTCTGAGAACAGTCATCAATGAGCCATCT 376

Qy       1332 TATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTGTAAAGCCCCTA 1391
              |||||||||||||| |||||| |||||||||| |||||||| || || || ||||||||
Db        377 TATAAAGAGAATGCCATGAGACTATCAAGAATCCACCATGACCAGCCAGTGAAGCCCCTG 436

Qy       1392 GATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCAAGCACCTGCGA 1451
              || |||||||||||||||| |||||||||||||| ||||||||||||||||||||| ||
Db        437 GACCGAGCAGTCTTCTGGATTGAGTTTGTCATGCGTCACAAAGGAGCCAAGCACCTTCGT 496

Qy       1452 TCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGATTGGGTTCCTG 1511
                || |||||||||||||| |||||| ||| || ||||| | ||||| ||||||||| |||
Db        497 GTGGCAGCCCATGACCTCAGCTGGTTTCAGTACCACTCTCTGGATGTAATTGGGTTTCTG 556

Qy       1512 CTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTATTTTCCTGTCAA 1571
              ||| |||||||||| |||||||||| ||| |  |||||||| ||| |||   | ||||
Db        557 CTGGCCTGTGTGGCATCTGCTATATTATTGGTTGCAAAATGTTGTTTGTTTATATTTCAA 616

Qy       1572 AAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAAATTCAAGAAAG 1631
              ||| ||  ||| ||  ||||||   ||||  | || ||| |      ||| ||    ||
Db        617 AAAGTTGGTAAGACAGGAAAGAAGAAAAAACGTGACTAGTTGAAGAAAAAGACACCTCAG 676

Qy       1632 ACCT 1635
                ||
Db        677 GTCT 680
```

RESULT 14
BQ942682
LOCUS BQ942682 938 bp mRNA linear EST 21-AUG-2002
DEFINITION AGENCOURT_8804675 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6337257 5', mRNA sequence.
ACCESSION BQ942682
VERSION BQ942682.1 GI:22358160
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 938)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Susan L. Sullivan, PhD.
cDNA Library Preparation: ResGen, Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation

```
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLAM13802  row: j  column: 10
          High quality sequence start: 26
          High quality sequence stop: 607.
FEATURES             Location/Qualifiers
     source          1. .938
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6337257"
                     /clone_lib="NIH_MGC_129"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: olfactory epithelium; Vector:
                     pCMV-SPORT6.1.ccdb; Site_1: EcoRV; Site_2: NotI; Cloned
                     unidirectionally. Primer: Oligo dT. Average insert size
                     2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this
                     is a NIH_MGC Library."
BASE COUNT      266 a    216 c    205 g    251 t
ORIGIN

  Query Match           28.3%;  Score 463.4;  DB 14;  Length 938;
  Best Local Similarity   80.7%;  Pred. No. 6.1e-102;
  Matches  553;  Conservative    0;  Mismatches  131;  Indels    1;  Gaps    1;

Qy        951 TCACTGTTTCAAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAG 1010
              || | |  ||||||  | ||||| |||||||| ||| ||||||| ||||||||||||||||
Db          3 TCTCGGGATCAAAACCTGACAGATGAAAAGGC-AATCTCATTGCCTCAGCCCTTGCCCAG 61

Qy       1011 ATCCCACAGAAGGTGTTATGGAGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAAT 1070
              || || |||||||| || ||| | |||||||||| | ||||  |||||||||| ||||| .
Db         62 ATTCCTCAGAAGGTTTTGTGGCGATACAAAGGAAAGATACCAGACACATTAGGATCCAAT 121

Qy       1071 ACTCGGCTGTATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCT 1130
              ||  | |||| ||||||||| || ||||||||||||||||| ||||||||||||| ||||
Db         122 ACAAGACTGTTTGATTGGATTCCTCAGAATGATCTTCTTGGACATCCCAAAACCAGAGCT 181

Qy       1131 TTTATCACTCATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATG 1190
              |||||||| |||||||||  ||||| |||||||| |||||||||||||| ||||||||
Db         182 TTTATCACCCATGGTGGAACAAATGGAATCTATGAGGCTATTTACCATGGGATCCCTATG 241

Qy       1191 GTGGGAGTTCCCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGA 1250
              |||||||| ||||| || | ||| |||| ||| ||||| ||||||||||||||||||||
Db         242 GTGGGAGTGCCCATGTTCGCTGACCAGCCTGACAACATTGCTCACATGAAGGCCAAAGGA 301

Qy       1251 GCAGCTGTAGAAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGA 1310
              ||||| || ||  | ||| | || || |||||||||  |||| | || |  ||| |||||
Db         302 GCAGCAGTGGAGGTGAACATGAACACGATGACAAGCTCAGATCTGCTCAATGCTCTGAGA 361

Qy       1311 ACAGTCATTACCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCAT 1370
              |||||||| |  ||  | |||||||||||||||||| |||||| |||||||||| ||||||
Db         362 ACAGTCATCAATGAGCCATCTTATAAAGAGAATGCCATGAGACTATCAAGAATCCACCAT 421

Qy       1371 GATCAACCTGTAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCAC 1430
              || || || ||||||||| || |||||||||||||||||| ||||||||||||||| |||
Db         422 GACCAGCCAGTGAAGCCCCTGGACCGAGCAGTCTTCTGGATTGAGTTTGTCATGCGTCAC 481

Qy       1431 AAAGGAGCCAAGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCT 1490
              |||||||||||||||| ||    || |||||||||||| |||||| ||| || |||||
Db         482 AAAGGAGCCAAGCACCTTCGTGTGGCAGCCCATGACCTCAGCTGGTTTCAGTACCACTCT 541

Qy       1491 ATAGATGTGATTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAA 1550
              | ||||| |||||||| |||||| |||||||||| |||||||||| ||| |  ||||
Db         542 CTGGATGTAATTGGGTTTCTGCTGGCCTGTGTGGCATCTGCTATATTATTGGTTGCAAAA 601

Qy       1551 TGTTTTTTATTTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAG 1610
              |||| ||| |||   | ||||||| ||  ||| ||  ||||||   ||||  | || |||
Db         602 TGTTGTTTGTTTATATTTCAAAAAGTTGGTAAGACAGGAAAGAAGAAAAAACGTGACTAG 661

Qy       1611 ATCTTTCCAAATTCAAGAAAGACCT 1635
              |     |||  ||    ||  ||
Db         662 TTGAAGAAAAAGACACCTCAGGTCT 686
```

```
RESULT 15
BG189133
LOCUS       BG189133                891 bp    mRNA    linear   EST 21-APR-2001
DEFINITION  RST8173 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION   BG189133
VERSION     BG189133.1  GI:13710820
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 891)
  AUTHORS   Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
            Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J.
            , Lerner,L., Costanzo,D., McElligott,K., Boozer,S., Mays,R., Smith
            ,E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K., Offenbacher
            ,J., Danzig,J. and Ducar,M.
  TITLE     Creation of genome-wide protein expression libraries using random
            activation of gene expression
  JOURNAL   Nat. Biotechnol. 19 (5), 440-445 (2001)
  MEDLINE   21227151
COMMENT     Contact: Scott J. Cain
            Athersys, Inc.
            3201 Carnegie Ave, Cleveland, OH 44115, USA
            Tel: 216 431 9900
            Fax: 216 361 9596
            Email: scain@athersys.com
            High quality sequence stop: 360.
FEATURES             Location/Qualifiers
     source          1. .891
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Athersys RAGE Library"
                     /cell_line="HT1080"
                     /note="See 'Creation of Genome-wide Protein Expression
                     Libraries using Random Activation of Gene Expression',
                     Nature Biotechnology, in press. Note that even though the
                     cell type indicated is HT1080, since a random activation
                     method was used, these sequence tags are not necessarily
                     expressed in HT1080 under normal circumstances."
BASE COUNT      268 a    167 c    208 g    244 t      4 others
ORIGIN

  Query Match           27.9%;  Score 456.4;  DB 12;  Length 891;
  Best Local Similarity   76.6%;  Pred. No. 3.1e-100;
  Matches  556;  Conservative    0;  Mismatches  170;  Indels    0;  Gaps    0;

Qy        892 AAATGGAAAATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGT 951
              |||||||  | ||||| ||||| || || ||| ||||| | |||||| ||||||||||||
Db          1 AAATGGAGGAGTTTGTACAGAGCTCTGGAGAAAATGGTGTGGTGGTGGTTTCTCTGGGGT 60

Qy        952 CACTGTTTCAAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGA 1011
              || || |    ||  | |||||||||| ||| ||  | |||||  |||||||||| |||
Db         61 CAATGGTCAGTAACATGACAGAAGAAAGGGCCAACGTAATTGCAACAGCCCTTGCCAAGA 120

Qy       1012 TCCCACAGAAGGTGTTATGGAGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATA 1071
              ||||||| |||||  | ||||| |    | || || ||||||    | |||||   |||||
Db         121 TCCCACAAAAGGTTCTTTGGAGATTTGATGGGAATAAACCAGATGCCTTAGGTCTCAATA 180

Qy       1072 CTCGGCTGTATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTT 1131
              |||| |||||  | |||||||.  ||||||| ||||| |||||||| ||||||| |||||
Db         181 CTCGACTGTACAAGTGGATACGGGAGAATGACCTTCTAGGTCATCCAAAAACCAGAGCTT 240

Qy       1132 TTATCACTCATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGG 1191
              |||| ||||||||||||   ||||| |||||||| || || ||||||||| |||||||||
Db         241 TTATAACTCATGGTGGAGCCAATGGCATCTATGAGGCAATCTACCATGGGATCCCTATGG 300

Qy       1192 TGGGAGTTCCCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAG 1251
              ||||  ||||  | |||  |||||| | |||||| || ||||||||||||||||| ||||
Db         301 TGGGCATTCCATTGTTTTTTGATCAACCTGATAATATTGCTCACATGAAGGCCAAGGGAG 360

Qy       1252 CAGCTGTAGAAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAA 1311
```

```
                ||||||| | | ||||||| || ||| | ||  ||| | |||| || ||| |
Db        361 CAGCTGTTAGAGTGGACTTCAACACAATGTCGAGTACAGACCTGCTGAATGCACTGAAGA 420

Qy       1312 CAGTCATTACCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATG 1371
                |||| |||| ||| | || |||||||||||| ||||| |||||||||||||| ||||
Db        421 CAGTAATTAATGATCCTTCATATAAAGAGAATATTATGAAATTATCAAGAATTCAACATG 480

Qy       1372 ATCAACCTGTAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACA 1431
                ||||||| || |||||||| |||||||||||||||||||| || ||||||||||||||||
Db        481 ATCAACCAGTGAAGCCCCTGGATCGAGCAGTCTTCTGGATTGAATTTGTCATGCGCCACA 540

Qy       1432 AAGGAGCCAAGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTA 1491
                 ||||||||| || || |||   || |||||  |||||||||| ||||| || |||||
Db        541 NAGGAGCCAAACATCTTCGAGTTGCAGCCCACAACCTCACCTGGNTCCAGTACCACTCTT 600

Qy       1492 TAGATGTGATTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAAT 1551
                | |||||||||||||||||||| | |||||||||| |   ||||  |  ||||| | |
Db        601 TGGATGTGATTGGGTTCCTGCTGGCTTGTGTGGCAACCGTGCTATTTATCATCACANAGT 660

Qy       1552 GTTTTTTATTTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGA 1611
                | | | | |||| |     || ||| || ||   |||||| | ||  ||||| |||
Db        661 GGTGTCTGTTTTGTTTCTGGAAGTTTGCTAGAAAAGGAAAGAAGGGAANAAGGGATTAGT 720

Qy       1612 TCTTTC 1617
                | | ||
Db        721 TATATC 726
```

Search completed: July 16, 2003, 14:55:49
Job time : 1792.01 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 16, 2003, 10:35:27 ; Search time 63.8246 Seconds
(without alignments)
7860.974 Million cell updates/sec

Title: US-09-981-353-165
Perfect score: 1636
Sequence: 1 gatcagtgtgtgagggaact.....ccaaattcaagaaagacctg 1636

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	775.6	47.4	2107	4	US-09-180-852-1	Sequence 1, Appli
	2	574	35.1	1413	4	US-09-813-918-1	Sequence 1, Appli
	3	329	20.1	2339	5	PCT-US92-00282-2	Sequence 2, Appli
	4	319.8	19.5	2336	5	PCT-US92-00282-1	Sequence 1, Appli
	5	145.6	8.9	391	4	US-09-370-838-21	Sequence 21, Appl
	6	97.4	6.0	1561	5	PCT-US92-00282-25	Sequence 25, Appl
	7	83.8	5.1	783	5	PCT-US92-00282-22	Sequence 22, Appl
	8	72.2	4.4	1008	5	PCT-US92-00282-10	Sequence 10, Appl
	9	62.4	3.8	1448	5	PCT-US92-00282-16	Sequence 16, Appl
	10	60	3.7	1066	5	PCT-US92-00282-14	Sequence 14, Appl
	11	55.4	3.4	1190	5	PCT-US92-00282-18	Sequence 18, Appl
	12	55	3.4	513	5	PCT-US92-00282-20	Sequence 20, Appl
	13	52.6	3.2	1197	5	PCT-US92-00282-12	Sequence 12, Appl
	14	50.6	3.1	1800	6	5180581-1	Patent No. 5180581
	15	50.6	3.1	2793	1	US-08-281-916-5	Sequence 5, Appli
	16	50.6	3.1	2793	2	US-08-460-725-7	Sequence 7, Appli
	17	48.8	3.0	1669	2	US-08-522-421-1	Sequence 1, Appli
	18	44	2.7	1627	3	US-09-106-464-1	Sequence 1, Appli
	19	41.4	2.5	51	5	PCT-US92-00282-37	Sequence 37, Appl
	20	38.8	2.4	1607	2	US-08-797-226-1	Sequence 1, Appli
c	21	38.6	2.4	2590	4	US-09-623-062-2	Sequence 2, Appli
	22	38.6	2.4	5194	1	US-09-599-652-1	Sequence 1, Appli
	23	38.6	2.4	5194	2	US-08-642-846-1	Sequence 1, Appli
	24	38.6	2.4	5194	4	US-09-264-604-1	Sequence 1, Appli
c	25	38.6	2.4	5333	4	US-09-623-062-1	Sequence 1, Appli
	26	38.2	2.3	2025	4	US-08-942-012B-23	Sequence 23, Appl
	27	37.2	2.3	1219	5	PCT-US92-00282-8	Sequence 8, Appli
c	28	36.2	2.2	2450	3	US-08-617-860B-9	Sequence 9, Appli
c	29	36	2.2	4342	4	US-09-338-907-107	Sequence 107, App
c	30	36	2.2	4342	4	US-09-218-207-107	Sequence 107, App
c	31	36	2.2	4582	4	US-09-338-907-118	Sequence 118, App
c	32	36	2.2	4582	4	US-09-218-207-118	Sequence 118, App
c	33	36	2.2	4686	4	US-09-338-907-117	Sequence 117, App
c	34	36	2.2	4686	4	US-09-218-207-117	Sequence 117, App
c	35	36	2.2	4875	4	US-09-338-907-114	Sequence 114, App
c	36	36	2.2	4875	4	US-09-218-207-114	Sequence 114, App
c	37	36	2.2	4958	4	US-09-338-907-116	Sequence 116, App
c	38	36	2.2	4958	4	US-09-218-207-116	Sequence 116, App
c	39	36	2.2	4986	4	US-09-338-907-121	Sequence 121, App
c	40	36	2.2	4986	4	US-09-218-207-121	Sequence 121, App
c	41	36	2.2	5020	4	US-09-338-907-120	Sequence 120, App
c	42	36	2.2	5020	4	US-09-218-207-120	Sequence 120, App
c	43	36	2.2	5044	4	US-09-338-907-115	Sequence 115, App
c	44	36	2.2	5044	4	US-09-218-207-115	Sequence 115, App
c	45	36	2.2	5057	4	US-09-338-907-123	Sequence 123, App

ALIGNMENTS

RESULT 1
US-09-180-852-1
; Sequence 1, Application US/09180852
; Patent No. 6287834
; GENERAL INFORMATION:
; APPLICANT: BELANGER, Alain
; APPLICANT: HUM, Dean W.
; APPLICANT: BEAULIEU, Martin
; APPLICANT: LEVESQUE, Eric
; TITLE OF INVENTION: CHARACTERIZATION AND USE OF AN ISOLATED URIDINE
; TITLE OF INVENTION: DIPHOSPHO-GLUCURONOSYLTRANSFERASE
; FILE REFERENCE: 1259-449
; CURRENT APPLICATION NUMBER: US/09/180,852
; CURRENT FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: PCT/CA97/00328
; EARLIER FILING DATE: 1997-05-16
; EARLIER APPLICATION NUMBER: US 08/649,319
; EARLIER FILING DATE: 1996-05-17
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2107
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (52)..(1644)
US-09-180-852-1

Query Match 47.4%; Score 775.6; DB 4; Length 2107;
Best Local Similarity 68.6%; Pred. No. 4.5e-216;
Matches 1104; Conservative 0; Mismatches 494; Indels 12; Gaps 2;

```
Qy          24 ATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGCTCCTGCAGCTCTTCTGT---GTT 80
               | || || ||||   || |     |  || ||||||||  ||||||||   |||    ||
Db          46 ACCAGGATGTCTCTGAAATGGATGTCAGTCTTTCTGCTGATGCAGCTCAGTTGTTACTTT 105

Qy          81 GGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTGACATGAGCCATTGGCTTAAT 140
               ||| |||    ||||| || || ||||||||||||   ||  | |||||||| | |||
Db         106 AGCTCTGGGAGTTGTGGAAAGGTGCTGGTGTGGCCCACAGAATACAGCCATTGGATAAAT 165

Qy         141 GTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATGAGGTAACAGTATTGACTCAC 200
               | |||   || || ||||||||  |   ||| || ||||||||  |  || |||||
Db         166 ATGAAGACAATCCTGGAAGAGCTTGTTCAGAGGGGTCATGAGGTGATTGTGTTGACATCT 225

Qy         201 TCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCATTGAAATTTGAGGTGGTCCAT 260
               ||    ||   |  | |  ||| ||  | |||||  | ||||| || ||    | |
Db         226 TCGGCTTCTATTCTTGTCAATGCCAGTAAATCATCTGCTATTAAATTAGAAGTTTATCCT 285

Qy         261 ATGCCACAGGACAGAACAGAAGAAAATGAAATATTT---------GTTGACCTAGCTCTG 311
```

```
Db        286 ACATCTTTAACTAAAAATGATTTGGAAGATTTTTTTATGAAAATGTTCGATAGATGGACA 345

Qy        312 AATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATGATTTTTTTGTT 371

Db        346 TATAGTATTTCAAAAAATACATTTTGGTCATATTTTTCACAACTACAAGAATTGTGTTGG 405

Qy        372 GAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATCAGACGCTTATG 431

Db        406 GAATATTCTGACTATAATATAAAGCTCTGTGAAGATGCAGTTTTGAACAAGAAACTTATG 465

Qy        432 AAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGATTCCCTGTGGA 491

Db        466 AGAAAACTACAAGAGTCAAAATTTGATGTCCTTCTGGCAGATGCCGTTAATCCCTGTGGT 525

Qy        492 GACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAATTTCTGTAGGA 551

Db        526 GAGCTGCTGGCTGAACTACTTAACATACCCTTTCTGTACAGTCTCCGCTTCTCTGTTGGC 585

Qy        552 GGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATGTACCTGTGCCT 611

Db        586 TACACAGTTGAGAAGAATGGTGGAGGATTTCTGTTCCCTCCTTCCTATGTACCTGTTGTT 645

Qy        612 ATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATTCAATGCTTTCA 671

Db        646 ATGTCAGAATTAAGTGATCAAATGATTTTCATGGAGAGGATAAAAAATATGATATATATG 705

Qy        672 GTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAGAGTTTTATAGT 731

Db        706 CTTTATTTTGACTTTTGGTTTCAAGCATATGATCTGAAGAAGTGGGACCAGTTTTATAGT 765

Qy        732 AAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTGAGATATGGCTA 791

Db        766 GAAGTTCTAGGAAGACCCACTACATTATTTGAGACAATGGGGAAAGCTGAAATGTGGCTC 825

Qy        792 ATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACTTTGAGTTTGTT 851

Db        826 ATTCGAACCTATTGGGATTTTGAATTTCCTCGCCCATTCTTACCAAATGTTGATTTTGTT 885

Qy        852 GGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAAATTTTGTCCAG 911

Db        886 GGAGGACTTCACTGTAAACCAGCCAAACCCTTGCCTAAGGAAATGGAAGAGTTTGTGCAG 945

Qy        912 AGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTCAAAATGTTACA 971

Db        946 AGCTCTGGAGAAAATGGTATTGTGGTGTTTTCTCTGGGGTCGATGATCAGTAACATGTCA 1005

Qy        972 GAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGAAGGTGTTATGG 1031

Db       1006 GAAGAAAGTGCCAACATGATTGCATCAGCCCTTGCCCAGATCCCACAAAAGGTTCTATGG 1065

Qy       1032 AGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGTATGATTGGATA 1091

Db       1066 AGATTTGATGGCAAGAAGCCAAATACTTTAGGTTCCAATACTCGACTGTATAAGTGGTTA 1125

Qy       1092 CCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTCATGGTGGAATG 1151

Db       1126 CCCCAGAATGACCTTCTTGGTCATCCCAAAACCAAAGCTTTTATAACTCATGGTGGAACC 1185

Qy       1152 AATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTCCCATATTTGGT 1211

Db       1186 AATGGCATCTATGAGGCGATCTACCATGGGATCCCTATGGTGGGCATTCCCTTGTTTGCG 1245

Qy       1212 GATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAGAAATAAACTTC 1271

Db       1246 GATCAACATGATAACATTGCTCACATGAAAGCCAAGGGAGCAGCCCTCAGTGTGGACATC 1305

Qy       1272 AAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTACCGATTCCTCT 1331

Db       1306 AGGACCATGTCAAGTAGAGATTTGCTCAATGCATTGAAGTCAGTCATTAATGACCCTATC 1365

Qy       1332 TATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTGTAAAGCCCCTA 1391
```

```
Db       1366 TATAAAGAGAATATCATGAAATTATCAAGAATTCATCATGATCAACCGGTGAAGCCCCTG 1425

Qy       1392 GATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCAAGCACCTGCGA 1451

Db       1426 GATCGAGCAGTCTTCTGGATTGAGTTTGTCATGCGCCATAAAGGAGCCAAGCACCTTCGG 1485

Qy       1452 TCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGATTGGGTTCCTG 1511

Db       1486 GTCGCAGCCCACAACCTCACCTGGATCCAGTACCACTCTTTGGATGTGATAGCATTCCTG 1545

Qy       1512 CTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTATTTTCCTGTCAA 1571

Db       1546 CTGGCCTGCGTGGCAACTATGATATTTATGATCACAAAATGTTGCCTGTTTTGTTTCCGA 1605

Qy       1572 AAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAAA 1621

Db       1606 AAGCTTGCCAAAACAGGAAAGAAGAAGAAAAGGGATTAGTTATATCAAAA 1655


RESULT 2
US-09-813-918-1
; Sequence 1, Application US/09813918
; Patent No. 6383789
; GENERAL INFORMATION:
;  APPLICANT: WEBSTER, Marion et al.
;  TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
;  TITLE OF INVENTION:  PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
;  TITLE OF INVENTION:  DRUG-METABOLIZING PROTEINS,
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: CL001175
;  CURRENT APPLICATION NUMBER: US/09/813,918
;  CURRENT FILING DATE:  2001-03-22
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 1413
;   TYPE: DNA
;   ORGANISM: Human
US-09-813-918-1

  Query Match             35.1%;  Score 574;  DB 4;  Length 1413;
  Best Local Similarity   78.4%;  Pred. No. 2e-157;
  Matches  688;  Conservative    0;  Mismatches  190;  Indels    0;  Gaps    0;

Qy        743 AAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTGAGATATGGCTAATACGAACATA 802

Db        520 ACGACCCACTACCTTATTTGAGACAATGGGAAAAGCTGACATATGGCTTATGCGAAACCC 579

Qy        803 TTGGGATTTTGAATTTCCTCAACCATACCAACCTAACTTTGAGTTTGTTGGAGGATTGCA 862

Db        580 CTGGAGTTTTCAATTTCCTCATCCATTCTTACCAAACGTTGATTTTGTTGGAGGATTCCA 639

Qy        863 CTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAAATTTTGTCCAGAGTTCAGGGGA 922

Db        640 CTGCAAACCTGCCAAACCCCTACCTAAGGAAATGGAGGAGTTTGTACAGAGCTCTGGAGA 699

Qy        923 AGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTCAAAATGTTACAGAAGAAAAGGC 982

Db        700 AAATGGTGTTGTGGTGTTTTCTCTGGGGTCAGTGATAAGTAACATGACAGCAGAAAGGGC 759

Qy        983 TAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGAAGGTGTTATGGAGGTACAAAGG 1042

Db        760 CAATGTAATTGCAACAGCCCTTGCCAGGATCCCACAAAAGGTTCTGTGGAGATTTGACGG 819

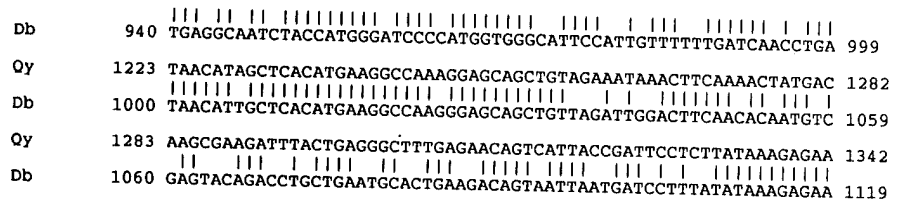
Qy       1043 AAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGTATGATTGGATACCCCAGAATGA 1102

Db        820 GAATAAACCAGATGCCTTAGGTCTCAATACTCGGCTGTACAAGTGGATACCCCAGAATGA 879

Qy       1103 TCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTCATGGTGGAATGAATGGGATCTA 1162

Db        880 CCTTCTAGGTCATCCAAAAACCAGAGCTTTTATAACTCATGGTGGAGCCAATGGCATCTA 939

Qy       1163 TGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTCCCATATTTGGTGATCAGCTTGA 1222
```



```
Db        940 TGAGGCAATCTACCATGGGATCCCCATGGTGGGCATTCCATTGTTTTTTGATCAACCTGA 999

Qy       1223 TAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAGAAATAAACTTCAAAACTATGAC 1282

Db       1000 TAACATTGCTCACATGAAGGCCAAGGGAGCAGCTGTTAGATTGGACTTCAACACAATGTC 1059

Qy       1283 AAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTACCGATTCCTCTTATAAAGAGAA 1342

Db       1060 GAGTACAGACCTGCTGAATGCACTGAAGACAGTAATTAATGATCCTTTATATAAAGAGAA 1119

Qy       1343 TGCTATGAGATTATCAAGAATTCACCATGATCAACCTGTAAAGCCCCTAGATCGAGCAGT 1402

Db       1120 TATTATGAAATTATCAAGAATTCAACATGATCAACCAGTAAAGCCCCTGGATCGAGCAGT 1179

Qy       1403 CTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCAAGCACCTGCGATCAGCTGCCCA 1462

Db       1180 CTTCTGGATTGAATTTGTCATGCCCCACAAAGGAGCCAAACACCTTCGAGTTGCAGCCCA 1239

Qy       1463 TGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGATTGGGTTCCTGCTGACCTGTGT 1522

Db       1240 TGACCTCACCTGGTTCCAGTACCACTCTTTGGATGTGATTGGGTTTCTGCTGGCCTGTGT 1299

Qy       1523 GGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTATTTTCCTGTCAAAAATTTAATAA 1582

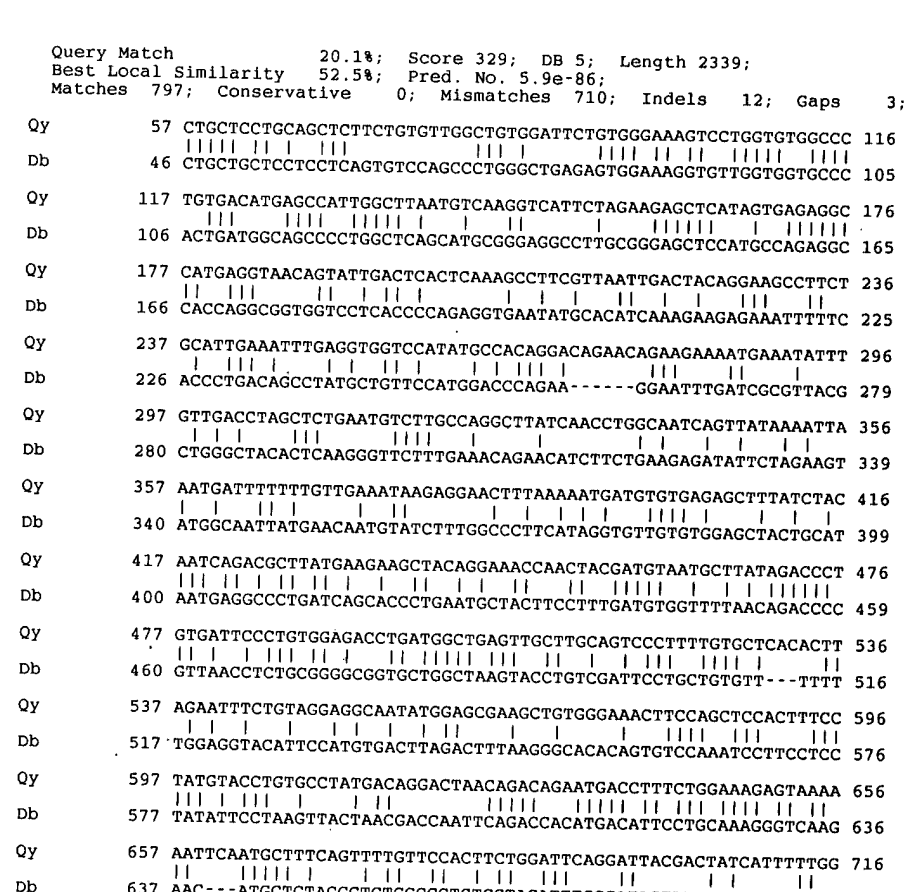
Db       1300 GGCAACTGTGATATTTATCATCACAAAGTTTTGTCTGTTTTGTTTCTGGAAGTTTGCTAG 1359

Qy       1583 AACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1620

Db       1360 AAAAGGGAAGAAGGGAAAAAGAGATTAGTTATGTCCGA 1397
```

RESULT 3
PCT-US92-00282-2
; Sequence 2, Application PC/TUS9200282
; GENERAL INFORMATION:
; APPLICANT: OWENS, IDA S.
; APPLICANT: RITTER, JOSEPH K.
; TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION
; TITLE OF INVENTION: THEREIN.
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN
; STREET: 1615 L STREET, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20036-5601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/00282
; FILING DATE: 19920110
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SCOTT, WATSON T.
; REGISTRATION NUMBER: 26581
; REFERENCE/DOCKET NUMBER: 91532-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3000
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2339 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
PCT-US92-00282-2

Query Match 20.1%; Score 329; DB 5; Length 2339;
Best Local Similarity 52.5%; Pred. No. 5.9e-86;
Matches 797; Conservative 0; Mismatches 710; Indels 12; Gaps 3;



```
Qy         57 CTGCTCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCC 116

Db         46 CTGCTGCTCCTCCTCAGTGTCCAGCCCTGGGCTGAGAGTGGAAAGGTGTTGGTGGTGCCC 105

Qy        117 TGTGACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGC 176

Db        106 ACTGATGGCAGCCCCTGGCTCAGCATGCGGGAGGCCTTGCGGGAGCTCCATGCCAGAGGC 165

Qy        177 CATGAGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCT 236

Db        166 CACCAGGCGGTGGTCCTCACCCCAGAGGTGAATATGCACATCAAAGAAGAGAAATTTTTC 225

Qy        237 GCATTGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTT 296

Db        226 ACCCTGACAGCCTATGCTGTTCCATGGACCCAGAA------GGAATTTGATCGCGTTACG 279

Qy        297 GTTGACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTA 356

Db        280 CTGGGCTACACTCAAGGGTTCTTTGAAACAGAACATCTTCTGAAGAGATATTCTAGAAGT 339

Qy        357 AATGATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTAC 416

Db        340 ATGGCAATTATGAACAATGTATCTTTGGCCCTTCATAGGTGTTGTGTGGAGCTACTGCAT 399

Qy        417 AATCAGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCT 476

Db        400 AATGAGGCCCTGATCAGCACCCTGAATGCTACTTCCTTTGATGTGGTTTTAACAGACCCC 459

Qy        477 GTGATTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTT 536

Db        460 GTTAACCTCTGCGGGGCGGTGCTGGCTAAGTACCTGTCGATTCCTGCTGTGTT---TTTT 516

Qy        537 AGAATTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCC 596

Db        517 TGGAGGTACATTCCATGTGACTTAGACTTTAAGGGCACACAGTGTCCAAATCCTTCCTCC 576

Qy        597 TATGTACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAA 656

Db        577 TATATTCCTAAGTTACTAACGACCAATTCAGACCACATGACATTCCTGCAAAGGGTCAAG 636

Qy        657 AATTCAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGG 716

Db        637 AAC---ATGCTCTACCCTCTGGCCCTGTCCTACATTTGCCATACTTTTTCTGCCCCTTAT 693

Qy        717 GAAGAGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAA 776

Db        694 GCAAGTCTTGCCTCTGAGCTTTTTCAGAGAGAGGTGTCAGTGGTGGATCTTGTCAGCTAT 753

Qy        777 GCTGAGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCT 836

Db        754 GCATCCGTGTGGCTGTTCCGAGGGGACTTTGTGATGGACTACCCCAGGCCGATCATGCCC 813

Qy        837 AACTTTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATG 896

Db        814 AACATGGTCTTCATTGGGGGCATCAACTGTGCCAACGGGAAGCCACTATCTCAGGAATTT 873

Qy        897 GAAAATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTG 956

Db        874 GAAGCCTACATTAATGCTTCTGGAGAACATGGAATTGTGGTTTTCTCTTTGGGATCAATG 933

Qy        957 TTTCAAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCA 1016

Db        934 GTCTCAGAAATTCCAGAGAAGAAAGCTATGGCAATTGCTGATGCTTTGGGCAAAATCCCT 993

Qy       1017 CAGAAGGTGTTATGGAGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGG 1076

Db        994 CAGACAGTCCTGTGGCGGTACACTGGAACCCGACCATCGAATCTTGCGAACAACACGATA 1053
```

```
Qy       1077 CTGTATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATC 1136
Db       1054 CTTGTTAAGTGGCTACCCCAAAACGATCTGCTTGGTCACCCGATGACCCGTGCCTTTATC 1113

Qy       1137 ACTCATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGA 1196
Db       1114 ACCCATGCTGGTTCCCATGGTGTTTATGAAAGCATATGCAATGGCGTTCCCATGGTGATG 1173

Qy       1197 GTTCCCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCT 1256
Db       1174 ATGCCCTTGTTTGGTGATCAGATGGACAATGCAAAGCGCATGGAGACTAAGGGAGCTGGA 1233

Qy       1257 GTAGAAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTC 1316
Db       1234 GTGACCCTGAATGTTCTGGAAATGACTTCTGAAGATTTAGAAAATGCTCTAAAAGCAGTC 1293

Qy       1317 ATTACCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAA 1376
Db       1294 ATCAATGACAAAAGTTACAAGGAGAACATCAGGCGCCTCTCCAGCCTTCACAAGGACCGC 1353

Qy       1377 CCTGTAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGA 1436
Db       1354 CCGGTGGAGCCGCTGGACCTGGCCGTGTTCTGGGTGGAGTTTGTGATGAGGCACAAGGGC 1413

Qy       1437 GCCAAGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGAT 1496
Db       1414 GCGCCACACCTGCGCCCCGCAGCCCACGACCTCACCTGGTACCAGTACCATTCCTTGGAC 1473

Qy       1497 GTGATTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTT 1556
Db       1474 GTGATTGGTTTCCTCTTGGCCGTCGTGCTGACAGTGGCCTTCATCACCTTTAAATGTTGT 1533

Qy       1557 TTATTTTCCTGTCAAAAAT 1575
Db       1534 GCTTATGGCTACCGGAAAT 1552


RESULT 4
PCT-US92-00282-1
; Sequence 1, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2336 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-1

  Query Match           19.5%;  Score 319.8;  DB 5;  Length 2336;
  Best Local Similarity   52.5%;  Pred. No. 2.8e-83;
  Matches  777;  Conservative    0;  Mismatches  692;  Indels   12;  Gaps    3;

Qy         95 TGGGAAAGTCCTGGTGTGGCCCTGTGACATGAGCCATTGGCTTAATGTCAAGGTCATTCT 154
Db         81 TGGGAAGATACTGTTGATCCCAGTGGATGGCAGCCACTGGCTGAGCATGCTTGGGGCCAT 140

Qy        155 AGAAGAGCTCATAGTGAGAGGCCATGAGGTAACAGTATTGACTCACTCAAAGCCTTCGTT 214
Db        141 CCAGCAGCTGCAGCAGAGGGGACATGAAATAGTTGTCCTAGCACCTGACGCCTCGTTGTA 200

Qy        215 AATTGACTACAGGAAGCCTTCTGCATTGAAATTTGAGGTGGTCCATATGCCACAGGACAG 274
Db        201 CATCAGAGACGGAGCATTTTACACCTTGA------AGACGTACCCTGTGCCATTCCAAAG 254

Qy        275 AACAGAAGAAAATGAAATATTTGTTGACCTAGCTCTGAATGTCTTGCCAGGCTTATCAAC 334
Db        255 GGAGGATGTGAAAGAGTCTTTTGTTAGTCTCGGGCATAATGTTTTTGAGAATGATTCTTT 314

Qy        335 CTGGCAATCAGTTATAAAATTAAATGATTTTTTTGTTGAAATAAGAGGAACTTTAAAAAT 394
Db        315 CCTGCAGCGTGTGATCAAAACATACAAGAAAATAAAAAAGGACTCTGCTATGCTTTTGTC 374

Qy        395 GATGTGTGAGAGCTTTATCTACAATCAGACGCTTATGAAGAAGCTACAGGAAACCAACTA 454
Db        375 TGGCTGTTCCCACTTACTGCACAACAAGGAGCTCATGGCCTCCCTGGCAGAAAGCAGCTT 434

Qy        455 CGATGTAATGCTTATAGACCCTGTGATTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGC 514
Db        435 TGATGTCATGCTGACGGACCCTTTCCTTCCTTGCAGCCCCATCGTGGCCCAGTACCTGTC 494

Qy        515 AGTCCCTTTTGTGCTCACACTTAGAATTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGG 574
Db        495 TCTGCCCACTGTATTC---TTCTTGCATGCACTGCCATGCAGCCTGGAATTTGAGGCTAC 551

Qy        575 GAAACTTCCAGCTCCACTTTCCTATGTACCTGTGCCTATGACAGGACTAACAGACAGAAT 634
Db        552 CCAGTGCCCCAACCCATTCTCCTACGTGCCCAGGCCTCTCTCCTCTCATTCAGATCACAT 611

Qy        635 GACCTTTCTGGAAAGAGTAAAAAATTCAATGCTTTCAGTTTTGTTCCACTTCTGGATTCA 694
Db        612 GACCTTCCTGCAGCGGGTGAAGAA---CATGCTCATTGCCTTTTCACAGAACTTTCTGTG 668

Qy        695 GGATTACGACTATCATTTTTGGGAAGAGTTTTATAGTAAGGCATTAGGAAGGCCCACTAC 754
Db        669 CGACGTGGTTTATTCCCCGTATGCAACCCTTGCCTCAGAATTCCTTCAGAGAGAGGTGAC 728

Qy        755 ATTATGTGAGACTGTGGGAAAAGCTGAGATATGGCTAATACGAACATATTGGGATTTTGA 814
Db        729 TGTCCAGGACCTATTGAGCTCTGCATCTGTCTGGCTGTTTAGAAGTGACTTTGTGAAGGA 788

Qy        815 ATTTCCTCAACCATACCAACCTAACTTTGAGTTTGTTGGAGGATTGCACTGTAAACCTGC 874
Db        789 TTACCCTAGGCCCATCATGCCCAATATGGTTTTTGTTGGTGGAATCAACTGCCTTCACCA 848

Qy        875 CAAAGCTTTGCCTAAGGAAATGGAAAATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGT 934
Db        849 AAATCCACTATCCCAGGAATTTGAAGCCTACATTAATGCTTCTGGAGAACATGGAATTGT 908

Qy        935 GGTGTTTTCTCTGGGGTCACTGTTTCAAAATGTTACAGAAGAAAAGGCTAATATCATTGC 994
Db        909 GGTTTTCTCTTTGGGATCAATGGTCTCAGAAATTCCAGAGAAGAAAGCTATGGCAATTGC 968

Qy        995 TTCAGCCCTTGCCCAGATCCCACAGAAGGTGTTATGGAGGTACAAAGGAAAAAAACCATC 1054
Db        969 TGATGCTTTGGGCAAAATCCCTCAGACAGTCCTGTGGCGGTACACTGGAACCCGACCATC 1028
```

```
Qy       1055 CACATTAGGAGCCAATACTCGGCTGTATGATTGGATACCCCAGAATGATCTTCTTGGTCA 1114
Db       1029 GAATCTTGCGAACAACACGATACTTGTTAAGTGGCTACCCCAAAACGATCTGCTTGGTCA 1088

Qy       1115 TCCCAAAACCAAAGCTTTTATCACTCATGGTGGAATGAATGGGATCTATGAAGCTATTTA 1174
Db       1089 CCCGATGACCCGTGCCTTTATCACCCATGCTGGTTCCCATGGTGTTTATGAAAGCATATG 1148

Qy       1175 CCATGGGGTCCCTATGGTGGGAGTTCCCATATTTGGTGATCAGCTTGATAACATAGCTCA 1234
Db       1149 CAATGGCGTTCCCATGGTGATGATGCCCTTGTTTGGTGATCAGATGGACAATGCAAAGCG 1208

Qy       1235 CATGAAGGCCAAAGGAGCAGCTGTAGAAATAAACTTCAAAACTATGACAAGCGAAGATTT 1294
Db       1209 CATGGAGACTAAGGGAGCTGGAGTGACCCTGAATGTTCTGGAAATGACTTCTGAAGATTT 1268

Qy       1295 ACTGAGGGCTTTGAGAACAGTCATTACCGATTCCTCTTATAAAGAGAATGCTATGAGATT 1354
Db       1269 AGAAAATGCTCTAAAAGCAGTCATCAATGACAAAAGTTACAAGGAGAACATCAGGCGCCT 1328

Qy       1355 ATCAAGAATTCACCATGATCAACCTGTAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGA 1414
Db       1329 CTCCAGCCTTCACAAGGACCGCCCGGTGGAGCCGCTGGACCTGGCCGTGTTCTGGGTGGA 1388

Qy       1415 GTTTGTCATGCGCCACAAAGGAGCCAAGCACCTGCGATCAGCTGCCCATGACCTCACCTG 1474
Db       1389 GTTTGTGATGAGGCACAAGGGCGCGCCACACCTGCGCCCCGCAGCCCACGACCTCACCTG 1448

Qy       1475 GTTCCAGCACTACTCTATAGATGTGATTGGGTTCCTGCTGACCTGTGTGGCAACTGCTAT 1534
Db       1449 GTACCAGTACCATTCCTTGGACGTGATTGGTTTCCTCTTGGCCGTCGTGCTGACAGTGGC 1508

Qy       1535 ATTCTTGTTCACAAAATGTTTTTTATTTTCCTGTCAAAAAT 1575
Db       1509 CTTCATCACCTTTAAATGTTGTGCTTATGGCTACCGGAAAT 1549
```

RESULT 5
US-09-370-838-21
; Sequence 21, Application US/09370838
; Patent No. 6444425
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Roadoh
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
; TITLE OF INVENTION: LUNG CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.475C1
; CURRENT APPLICATION NUMBER: US/09/370,838
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/285,323
; EARLIER FILING DATE: 1999-04-02
; NUMBER OF SEQ ID NOS: 289
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 21
; LENGTH: 391
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-370-838-21

Query Match 8.9%; Score 145.6; DB 4; Length 391;
Best Local Similarity 61.2%; Pred. No. 6.4e-33;
Matches 235; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

```
Qy        925 ATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTCAAAATGTTACAGAAGAAAAGGCTA 984
Db          1 ATGGAATTGTGGTTTTCTCTTTGGGATCAATGGTCTCAGAAATTCCAGAGAAGAAAGCTG 60

Qy        985 ATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGAAGGTGTTATGGAGGTACAAAGGAA 1044
Db         61 TGGCGATTGCTGATGCTTTGGGCAAAATCCCTCAGACAGTCCTGTGGCGGTACACTGGAA 120

Qy       1045 AAAAACCATCCACATTAGGAGCCAATACTCGGCTGTATGATTGGATACCCCAGAATGATC 1104
Db        121 CCCGACCATCGAATCTTGCGAACAACACGATACTTGTTCAGTGGCTACCCCAAAACGATC 180

Qy       1105 TTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTCATGGTGGAATGAATGGGATCTATG 1164
Db        181 TGCTTGGTCACCCAATGACCCGTGCCTTTATCACCCATGCTAGTTCCCATGGTGTTAATG 240

Qy       1165 AAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTCCCATATTTGGTGATCAGCTTGATA 1224
Db        241 AAAGCATATGCAATGGCGTTCCCATGGTGATGATACCCTTATTTGGTGATCAGATGGACA 300

Qy       1225 ACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAGAAATAAACTTCAAAACTATGACAA 1284
Db        301 ATGCAAAGCGCAGGGAGACTAAGGGAGCTGGAGTGACCCTGAATGTTCTGGAGATGACTT 360

Qy       1285 GCGAAGATTTACTGAGGGCTTTGA 1308
Db        361 CTGAAGATCTAGAAGATGCTCTGA 384
```

RESULT 6
PCT-US92-00282-25
; Sequence 25, Application PC/TUS9200282
; GENERAL INFORMATION:
; APPLICANT: OWENS, IDA S.
; APPLICANT: RITTER, JOSEPH K.
; TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION
; TITLE OF INVENTION: THEREIN.
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN
; STREET: 1615 L STREET, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20036-5601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/00282
; FILING DATE: 19920110
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SCOTT, WATSON T.
; REGISTRATION NUMBER: 26581
; REFERENCE/DOCKET NUMBER: 91532-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3000
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1561 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
PCT-US92-00282-25

Query Match 6.0%; Score 97.4; DB 5; Length 1561;
Best Local Similarity 70.1%; Pred. No. 1.4e-18;
Matches 131; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

```
Qy       1331 TTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTGTAAAGCCCCT 1390
Db        271 TTACAAGGAGAACATCATGCGCCTCTCCAGCCTTCACAAGGACCGCCCGGTGGAGCCGCT 330

Qy       1391 AGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCAAGCACCTGCG 1450
```

```
Db        331 GGACCTGGCCGTGTTCTGGGTGGAGTTTGTGATGAGGCACAAGGGCGCGCCACACCTGCG 390

Qy       1451 ATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGATTGGGTTCCT 1510

Db        391 CCCCGCAGCCCACGACCTCACCTGGTACCAGTACCATTCCTTGGACGTGATTGGTTTCCT 450

Qy       1511 GCTGACC 1517

Db        451 CTTGGCC 457


RESULT 7
PCT-US92-00282-22
; Sequence 22, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  22:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  783 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-22

  Query Match             5.1%;  Score 83.8;  DB 5;  Length 783;
  Best Local Similarity   61.9%;  Pred. No. 9.1e-15;
  Matches  133;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy       1110 GGTCATCCCAAAACCAAAGCTTTTATCACTCATGGTGGAATGAATGGGATCTATGAAGCT 1169

Db        471 GGTCACCCGATGACCCGTGCCTTTATCACCCATGCTGGTTCCCATGGTGTTTATGAAAGC 530

Qy       1170 ATTTACCATGGGGTCCCTATGGTGGGAGTTCCCATATTTGGTGATCAGCTTGATAACATA 1229

Db        531 ATATGCAATGGCGTTCCCATGGTGATGATGCCCTTGTTTGGTGATCAGATGGACAATGCA 590

Qy       1230 GCTCACATGAAGGCCAAAGGAGCAGCTGTAGAAATAAACTTCAAAACTATGACAAGCGAA 1289

Db        591 AAGCGCATGGAGACTAAGGGAGCTGGAGTGACCCTGAATGTTCTGGAAATGACTTCTGAA 650

Qy       1290 GATTTACTGAGGGCTTTGAGAACAGTCATTACCGA 1324

Db        651 GATTTAGAAAATGCTCTAAAAGCAGTCATCAATGA 685


RESULT 8
PCT-US92-00282-10
; Sequence 10, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1008 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-10

  Query Match             4.4%;  Score 72.2;  DB 5;  Length 1008;
  Best Local Similarity   45.5%;  Pred. No. 2.5e-11;
  Matches  380;  Conservative    0;  Mismatches  443;  Indels   12;  Gaps    3;

Qy         57 CTGCTCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCC 116

Db        157 CTGCTGCTTCTCCTCAGTGTCCAGCCCTGGGCTGAGAGTGGGAAGGTGCTGGTGGTGCCC 216

Qy        117 TGTGACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGC 176

Db        217 ACTGATGGCAGCCACTGGCTCAGCATGCGGGAGGCCTTGCGGGACCTCCATGCGAGAGGC 276

Qy        177 CATGAGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCT 236

Db        277 CACCAGGTGGTGGTCCTCACCCTGGAGGTGAATATGTACATCAAAGAAGAGAACTTTTTC 336

Qy        237 GCATTGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTT 296

Db        337 ACCCTGACAACGTATGCCATTTCATGGACCCAGGAC------GAATTTGATCGCCTTTTG 390

Qy        297 GTTGACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTA 356

Db        391 CTGGGTCACACTCAATCGTTCTTTGAAACAGAACATCTTCTGATGAAATTTTCTAGAAGA 450

Qy        357 AATGATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTAC 416

Db        451 ATGGCAATTATGAACAATATGTCTTTGATCATACATAGGTCTTGTGTGGAGCTACTGCAT 510

Qy        417 AATCAGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCT 476
```

```
Db        511 AATGAGGCCCTGATCAGGCACCTGCATGCTACTTCCTTTGATGTGGTTCTAACAGACCCC 570

Qy        477 GTGATTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTT 536

Db        571 TTTCACCTCTGCGCGGCGGTGCTGGCTAAGTACCTGTCGATTCCTGCTGTGTTTTTCTTG 630

Qy        537 AGAATTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCC 596

Db        631 AGGAACATTCCATGTG---ATTTAGACTTTAAGGGCACACAGTGTCCAAACCCTTCCTCC 687

Qy        597 TATGTACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAA 656

Db        688 TATATTCCTAGATTACTAACGACCAATTCAGACCACATGACATTCCTGCAAAGGGTCAAG 747

Qy        657 AATTCAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGG 716

Db        748 AA---CATGCTCTACCCTCTGGCCCTGTCCTACCTTTGCCATGCTGTTTCTGCTCCTTAT 804

Qy        717 GAAGAGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAA 776

Db        805 GCAAGCCTTGCCTCTGAGCTTTTTCAGAGAGAGGTGTCAGTGGTGGATCTTGTCAGCCAT 864

Qy        777 GCTGAGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCT 836

Db        865 GCATCCTGGTGGCTGTTCCGAGGGGAGTTTGTGATGGATTACCCCAGGCCGATCATGCCC 924

Qy        837 AACTTTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGG 891

Db        925 AACATGGTCTTCATTGGGGGCATCAACTGTGCCAACGGGAAGCCACTATCTCAGG 979
```

RESULT 9
PCT-US92-00282-16
; Sequence 16, Application PC/TUS9200282
; GENERAL INFORMATION:
; APPLICANT: OWENS, IDA S.
; APPLICANT: RITTER, JOSEPH K.
; TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION
; TITLE OF INVENTION: THEREIN.
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN
; STREET: 1615 L STREET, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20036-5601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/00282
; FILING DATE: 19920110
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SCOTT, WATSON T.
; REGISTRATION NUMBER: 26581
; REFERENCE/DOCKET NUMBER: 91532-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3000
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1448 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: YES

PCT-US92-00282-16

Query Match 3.8%; Score 62.4; DB 5; Length 1448;
Best Local Similarity 48.5%; Pred. No. 2.1e-08;
Matches 233; Conservative 0; Mismatches 241; Indels 6; Gaps 2;

```
Qy        387 TTAAAAATGATGTGTGAGAGCTTTATCTACAATCAGACGCTTATGAAGAAGCTACAGGAA 446

Db        475 TTTCATAGGTCTTGTGTGGAGCTACTGCATAATGAGGCCCTGATCAGGCACCTGAATGCT 534

Qy        447 ACCAACTACGATGTAATGCTTATAGACCCTGTGATTCCCTGTGGAGACCTGATGGCTGAG 506

Db        535 ACTTCCTTCGATGTGGTTTTAATAGACCCAATTTACCTCTGTGGGGCAGTGCTGGCTAAG 594

Qy        507 TTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAATTTCTGTAGGAGGCAATATGGAGCGA 566

Db        595 TACCTATCAATTCCTGCTGTGTTTTTTTTGAGGAACATTCCATGTG---ATTTAGACTTT 651

Qy        567 AGCTGTGGGAAACTTCCAGCTCCACTTTCCTATGTACCTGTGCCTATGACAGGACTAACA 626

Db        652 AAGGGCACACAGTGTCCAAATCCTTCCTCCTATATTCCTAAGTTACTAACGACCAATTCA 711

Qy        627 GACAGAATGACCTTTCTGGAAAGAGTAAAAAATTCAATGCTTTCAGTTTTGTTCCACTTC 686

Db        712 GACCACATGACATTCCTGCAAAGGGTCAAGAA---CATGCTCTACCCTCTGGCCCTGTCC 768

Qy        687 TGGATTCAGGATTACGACTATCATTTTTGGGAAGAGTTTTATAGTAAGGCATTAGGAAGG 746

Db        769 TACCTTTGCCATGCTCTTTCTGCTCCTTATGCAAGCCTTGCCTCTGAGCTTTTTCAGAGG 828

Qy        747 CCCACTACATTATGTGAGACTGTGGGAAAAGCTGAGATATGGCTAATACGAACATATTGG 806

Db        829 GAGGTGTCAGTGGTGGATCTTCTCAGCCATGCATCTGTGTGGCTGTTCCGAGGGGACTTT 888

Qy        807 GATTTTGAATTTCCTCAACCATACCAACCTAACTTTGAGTTTGTTGGAGGATTGCACTGT 866

Db        889 GTGATGGATTACCCCAGGCCGATCATGCCCAACATGGTCTTCATTGGGGGCATCAACTGT 948
```

RESULT 10
PCT-US92-00282-14
; Sequence 14, Application PC/TUS9200282
; GENERAL INFORMATION:
; APPLICANT: OWENS, IDA S.
; APPLICANT: RITTER, JOSEPH K.
; TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION
; TITLE OF INVENTION: THEREIN.
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN
; STREET: 1615 L STREET, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20036-5601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/00282
; FILING DATE: 19920110
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SCOTT, WATSON T.
; REGISTRATION NUMBER: 26581
; REFERENCE/DOCKET NUMBER: 91532-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3000
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 14:

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1066 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-14

  Query Match             3.7%;  Score 60;  DB 5;  Length 1066;
  Best Local Similarity   44.9%;  Pred. No. 9.1e-08;
  Matches  360;  Conservative    0;  Mismatches  430;  Indels   12;  Gaps    3;

Qy         57 CTGCTCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCC 116
Db        176 CTGCTGCTTCTCCTCAGTGTCCAGCCCTGGGCTGAGAGTGGAAAGGTGTTGGTGGTGCCC 235

Qy        117 TGTGACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGC 176
Db        236 ATTCATGGCAGCCACTGGCTCAGCATGCGGGAGGTCTTGCGGGAGCTCCATGCCAGAGGC 295

Qy        177 CATGAGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCT 236
Db        296 CACCAGGCAGTGGTCCTCACCCCAGAGGTGAATATGCACATCAAAGAAGAGAACTTTTTC 355

Qy        237 GCATTGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTT 296
Db        356 ACCCTGACAACCTATGCCATTTCGTGGACCCAGGATGAATTTGATCGCCATG------TG 409

Qy        297 GTTGACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTA 356
Db        410 CTGGGCCACACTCAACTGTACTTTGAAACAGAACATTTTCTGAAGAAATTTTTCAGAAGT 469

Qy        357 AATGATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTAC 416
Db        470 ATGGCAATGTTGAACAATATGTCTTTGGTCTATCATAGGTCTTGTGTGGAGCTACTACAT 529

Qy        417 AATCAGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCT 476
Db        530 AATGAGGCCCTGATCAGGCACCTGAATGCTACTTCCTTTGATGTGGTTTTAACAGACCCC 589

Qy        477 GTGATTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTT 536
Db        590 GTTAACCTCTGCGCGGCAGTGCTGGCTAAGTACCTGTCGATTCCTACTGTGTTTTTTTTG 649

Qy        537 AGAATTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCC 596
Db        650 AGGAACATTCCATGTG---ATTTAGACTTTAAGGGCACACAGTGTCCAAACCCTTCCTCC 706

Qy        597 TATGTACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAA 656
Db        707 TATATTCCTAGATTACTAACAACCAATTCAGACCACATGACATTCATGCAAAGGGTCAAG 766

Qy        657 AATTCAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGG 716
Db        767 AAC---ATGCTCTACCCTCTGGCCCTGTCCTACATTTGCCATGCTTTTTCTGCTCCTTAT 823

Qy        717 GAAGAGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAA 776
Db        824 GCCAGCCTTGCCTCTGAGCTTTTTCAGAGAGAGGTGTCAGTGGTGGATATTCTCAGTCAT 883

Qy        777 GCTGAGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCT 836
Db        884 GCATCTGTGTGGCTGTTCCGAGGGGACTTTGTGATGGACTACCCCAGGCCAATCATGCCC 943

Qy        837 AACTTTGAGTTTGTTGGAGGAT 858
Db        944 AACATGGTCTTCATTGGGGCAT 965


RESULT 11
PCT-US92-00282-18
; Sequence 18, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1190 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-18

  Query Match             3.4%;  Score 55.4;  DB 5;  Length 1190;
  Best Local Similarity   45.9%;  Pred. No. 2.1e-06;
  Matches  271;  Conservative    0;  Mismatches  311;  Indels    9;  Gaps    2;

Qy         95 TGGGAAAGTCCTGGTGTGGCCCTGTGACATGAGCCATTGGCTTAATGTCAAGGTCATTCT 154
Db        165 TGGGAAGATACTGTTGATCCCAGTGGATGGCAGCCACTGGCTGAGCATGCTTGGGGCCAT 224

Qy        155 AGAAGAGCTCATAGTGAGAGGCCATGAGGTAACAGTATTGACTCACTCAAAGCCTTCGTT 214
Db        225 CCAGCAGCTGCAGCAGAGGGGACATGAAATAGTTGTCCTAGCACCTGACGCCTCGTTGTA 284

Qy        215 AATTGACTACAGGAAGCCTTCTGCATTGAAATTTGAGGTGGTCCATATGCCACAGGACAG 274
Db        285 CATCAGAGACGGAGCATTTTACACCTTGA------AGACGTACCCTGTGCCAATCCAAAG 338

Qy        275 AACAGAAGAAAATGAAATATTTGTTGACCTAGCTCTGAATGTCTTGCCAGGCTTATCAAC 334
Db        339 GGAGGATGTGAAAGAGTCTTTTGTTAGTCTCGGGCATAATGTTTTTGAGAATGATTCTTT 398

Qy        335 CTGGCAATCAGTTATAAAATTAAATGATTTTTTGTTGAAATAAGAGGAACTTTAAAAAT 394
Db        399 CCTGCAGCGTGTGATCAAAACATACAAGAAAATAAAAAAGGACTCTGCTATGCTTTTGTC 458

Qy        395 GATGTGTGAGAGCTTTATCTACAATCAGACGCTTATGAAGAAGCTACAGGAAACCAACTA 454
Db        459 TGGCTGTTCCCACTTACTGCACAACAAGGAGCTCATGGCCTCCCTGGCAGAAAGCAGCTT 518

Qy        455 CGATGTAATGCTTATAGACCCTGTGATTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGC 514
Db        519 TGATGTCATGCTGACGGACCCTTTCCTTCCTTGCAGCCCCATCGTGGCCCAGTACCTGTC 578

Qy        515 AGTCCCTTTTGTGCTCACACTTAGAATTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGG 574
Db        579 TCTGCCCACTGTATTCTTCTTGC---ATGCACTGCCATGCAGCCTGGAATTTGAGGCTAC 635
```

```
Qy        575 GAAACTTCCAGCTCCACTTTCCTATGTACCTGTGCCTATGACAGGACTAACAGACAGAAT 634
Db        636 CCAGTGCCCCAACCCATTCTCCTACGTGCCCAGGCCTCTCTCCTCTCATTCAGATCACAT 695

Qy        635 GACCTTTCTGGAAAGAGTAAAAAATTCAATGCTTTCAGTTTTGTTCCACTT 685
Db        696 GACCTTCCTGCAGCGGGTGAAGAACATGCTCATTGCCTTTTCACAGAACTT 746


RESULT 12
PCT-US92-00282-20
; Sequence 20, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  513 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-20

  Query Match             3.4%;  Score 55;  DB 5;  Length 513;
  Best Local Similarity   61.5%;  Pred. No. 1.8e-06;
  Matches   88;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy        882 TTGCCTAAGGAAATGGAAAATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTT 941
Db        222 TGGCTCTAGGAATTTGAAGCCTACATTAATGCTTCTGGAGAACATGGAATTGTGGTTTTC 281

Qy        942 TCTCTGGGGTCACTGTTTCAAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCC 1001
Db        282 TCTTTGGGATCAATGGTCTCAGAAATTCCAGAGAAGAAAGCTATGGCAATTGCTGATGCT 341

Qy       1002 CTTGCCCAGATCCCACAGAAGGT 1024
Db        342 TTGGGCAAAATCCCTCAGACAGT 364


RESULT 13
PCT-US92-00282-12
; Sequence 12, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1197 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-12

  Query Match             3.2%;  Score 52.6;  DB 5;  Length 1197;
  Best Local Similarity   47.1%;  Pred. No. 1.4e-05;
  Matches  232;  Conservative    0;  Mismatches  254;  Indels    7;  Gaps    2;

Qy        399 TGTGAGAGCTTTATCTACAATCAGACGCTTATGAAGAAGCTACAGGAAACCAACTACGAT 458
Db        612 TGTGTGGAGCTACTGCATAATGAGGCCCTGATCAGGCACCTGAATGCTACTTCCTTTGAT 671

Qy        459 GTAATGCTTATAGACCCTGTGATTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTC 518
Db        672 GTGGTTTTAACAGACCCCGTTAACCTCTGTGGGGCGGTGCTGGCTAAGTACCTGTCGATT 731

Qy        519 CCTTTTGTGCTCACACTTAGAATTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAA 578
Db        732 CCTGCTGTGTTTTTTTGGAGGTACATTCCATGTGAC----TTAGACTTTAAGGCCATACA 787

Qy        579 CTTCCAGCTCCACTTTCCTATGTACCTGTGCCTATGACAGGACTAACAGACAGAATGACC 638
Db        788 GTGTCAAATCCTTCCTCCTATATTCCTAAGTTACTAACGACCAATTCAGACCACATGACA 847

Qy        639 TTTCTGGAAAGAGTAAAAAATTCAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGAT 698
Db        848 TTCCTGCAAAGGGTCAAGAAC---ATGCTCTACCCTCTGGCCCTGTCCTACATTTGCCAT 904

Qy        699 TACGACTATCATTTTTGGGAAGAGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTA 758
Db        905 ACTTTTTCTGCCCCTTATGCAAGTCTTGCCTCTGAGCTTTTTCAGAGAGAGGTGTCAGTG 964

Qy        759 TGTGAGACTGTGGGAAAAGCTGAGATATGGCTAATACGAACATATTGGGATTTTGAATTT 818
Db        965 GTGGATCTTGTCAGCTATGCATCCGTGTGGCTGTTCCGAGGGGACTTTGTGATGGACTAC 1024

Qy        819 CCTCAACCATACCAACCTAACTTTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAA 878
Db       1025 CCCAGGCCGATCATGCCCAACATGGTCTTCATTGGGGGCATCAACTGTGCCAACGGGAAG 1084
```

```
Qy        879 GCTTTGCCTAAGG 891
                |  |  || |||
Db       1085 CCACTATCTCAGG 1097


RESULT 14
5180581-1
;Patent No. 5180581
;     APPLICANT: MILLER, LOIS K.;O'REILLY, DAVID R.
;     TITLE OF INVENTION: BIOLOGICAL INSECT CONTROL AGENTS AND
;METHODS OF USE
;     NUMBER OF SEQUENCES: 2
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/373,952
;       FILING DATE: 29-JUN-1989
;SEQ ID NO:1:
;       LENGTH: 1800
5180581-1

  Query Match             3.1%;  Score 50.6;  DB 6;  Length 1800;
  Best Local Similarity   49.1%;  Pred. No. 6.5e-05;
  Matches  134;  Conservative    0;  Mismatches  139;  Indels    0;  Gaps    0;

Qy       1036 ACAAAGGAAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGTATGATTGGATACCCC 1095
              || ||| |  ||||      || ||    |||||         | |  ||||| |    |
Db       1143 ACGAAGTAGTAAAAAACATAACGTTGCCCGCCAACGTAATCACGCAAAATTGGTTTAATC 1202

Qy       1096 AGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTCATGGTGGAATGAATG 1155
              |    |   | ||  |||||   ||||     || ||||| || || || ||| |  |
Db       1203 AACGCGCCGTGCTGCGTCATAAAAAAATGGCGGCGTTTATTACGCAAGGCGGACTACAAT 1262

Qy       1156 GGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTCCCATATTTGGTGATC 1215
               || | | || ||  |  |    ||| | || ||||||  |  | |||||  | || || |
Db       1263 CGAGCGACGAGGCCTTGGAAGCCGGGATACCCATGGTGTGTCTGCCCATGATGGGCGACC 1322

Qy       1216 AGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAGAAATAAACTTCAAAA 1275
              || || |  |        |  |  ||      || | |||    |   |  ||      |
Db       1323 AGTTTTACCATGCGCACAAATTACAGCAACTCGGCGTAGCCCGCGCCTTGGACACTGTTA 1382

Qy       1276 CTATGACAAGCGAAGATTTACTGAGGGCTTTGA 1308
              |  |  | |||||  |  ||||   |||  | |
Db       1383 CCGTTTCCAGCGATCAACTACTAGTGGCGATAA 1415


RESULT 15
US-08-281-916-5
; Sequence 5, Application US/08281916
; Patent No. 5662897
;  GENERAL INFORMATION:
;     APPLICANT:  Miller, Lois K.
;     APPLICANT:  Black, Bruce C.
;     APPLICANT:  Dierks, Peter M.
;     APPLICANT:  Fleming, Nancy C.
;     APPLICANT:  Ahmed, Fakhruddin
;     TITLE OF INVENTION:  Insect Viruses, Sequences, Insecticidal
;     TITLE OF INVENTION:  Compositions and Methods of Use
;     NUMBER OF SEQUENCES:  8
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Greenlee and Winner, P.C.
;       STREET:  5370 Manhattan Circle, Suite 201
;       CITY:  Boulder
;       STATE:  Colorado
;       COUNTRY:  USA
;       ZIP:  80303
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/281,916
;       FILING DATE:  27-JUL-1994
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Ferber, Donna M.
;       REGISTRATION NUMBER:  33,878
;       REFERENCE/DOCKET NUMBER:  29-94
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (303) 499-8080
;       TELEFAX:  (303) 499-8089
;       TELEX:  49617824
;  INFORMATION FOR SEQ ID NO:  5:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2793 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
US-08-281-916-5

  Query Match             3.1%;  Score 50.6;  DB 1;  Length 2793;
  Best Local Similarity   49.1%;  Pred. No. 8.1e-05;
  Matches  134;  Conservative    0;  Mismatches  139;  Indels    0;  Gaps    0;

Qy       1036 ACAAAGGAAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGTATGATTGGATACCCC 1095
              || ||| |  ||||      || ||    |||||         | |  ||||| |    |
Db       1143 ACGAAGTAGTAAAAAACATAACGTTGCCCGCCAACGTAATCACGCAAAATTGGTTTAATC 1202

Qy       1096 AGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTCATGGTGGAATGAATG 1155
              |    |   | ||  |||||   ||||     || ||||| || || || ||| |  |
Db       1203 AACGCGCCGTGCTGCGTCATAAAAAAATGGCGGCGTTTATTACGCAAGGCGGACTACAAT 1262

Qy       1156 GGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTCCCATATTTGGTGATC 1215
               || | | || ||  |  |    ||| | || ||||||  |  | |||||  | || || |
Db       1263 CGAGCGACGAGGCCTTGGAAGCCGGGATACCCATGGTGTGTCTGCCCATGATGGGCGACC 1322

Qy       1216 AGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAGAAATAAACTTCAAAA 1275
              || || |  |        |  |  ||      || | |||    |   |  ||      |
Db       1323 AGTTTTACCATGCGCACAAATTACAGCAACTCGGCGTAGCCCGCGCCTTGGACACTGTTA 1382

Qy       1276 CTATGACAAGCGAAGATTTACTGAGGGCTTTGA 1308
              |  |  | |||||  |  ||||   |||  | |
Db       1383 CCGTTTCCAGCGATCAACTACTAGTGGCGATAA 1415


Search completed: July 16, 2003, 14:58:55
Job time : 67.8246 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 16, 2003, 13:28:07 ; Search time 215.273 Seconds
(without alignments)
12133.023 Million cell updates/sec

Title: US-09-981-353-165
Perfect score: 1636
Sequence: 1 gatcagtgtgtgagggaact.....ccaaattcaagaaagacctg 1636

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1114047 seqs, 798260406 residues

Total number of hits satisfying chosen parameters: 2228094

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
6: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1636	100.0	1636	9	US-09-981-353-165	Sequence 165, App
2	1636	100.0	2966	9	US-09-981-353-33	Sequence 33, Appl
3	1636	100.0	2974	9	US-10-174-590-521	Sequence 521, App
4	1636	100.0	2974	9	US-10-176-758-521	Sequence 521, App
5	1636	100.0	2974	9	US-10-175-737-521	Sequence 521, App
6	1636	100.0	2974	9	US-10-173-706-521	Sequence 521, App
7	1636	100.0	2974	9	US-10-175-738-521	Sequence 521, App
8	1636	100.0	2974	9	US-10-175-752-521	Sequence 521, App
9	1636	100.0	2974	9	US-10-176-482-521	Sequence 521, App
10	1636	100.0	2974	9	US-10-176-757-521	Sequence 521, App
11	1636	100.0	2974	9	US-10-176-913-521	Sequence 521, App
12	1636	100.0	2974	9	US-10-180-552-521	Sequence 521, App
13	1636	100.0	2974	9	US-10-180-557-521	Sequence 521, App
14	1636	100.0	2974	9	US-10-173-700-521	Sequence 521, App
15	1636	100.0	2974	9	US-10-174-572-521	Sequence 521, App
16	1636	100.0	2974	9	US-10-174-579-521	Sequence 521, App
17	1636	100.0	2974	9	US-10-174-582-521	Sequence 521, App
18	1636	100.0	2974	9	US-10-174-588-521	Sequence 521, App
19	1636	100.0	2974	9	US-10-175-739-521	Sequence 521, App
20	1636	100.0	2974	9	US-10-175-740-521	Sequence 521, App
21	1636	100.0	2974	9	US-10-175-743-521	Sequence 521, App
22	1636	100.0	2974	9	US-10-176-488-521	Sequence 521, App
23	1636	100.0	2974	9	US-10-176-492-521	Sequence 521, App
24	1636	100.0	2974	9	US-10-176-747-521	Sequence 521, App
25	1636	100.0	2974	9	US-10-176-750-521	Sequence 521, App
26	1636	100.0	2974	9	US-10-176-985-521	Sequence 521, App
27	1636	100.0	2974	9	US-10-176-987-521	Sequence 521, App
28	1636	100.0	2974	9	US-10-176-991-521	Sequence 521, App
29	1636	100.0	2974	9	US-10-176-992-521	Sequence 521, App
30	1636	100.0	2974	9	US-10-176-993-521	Sequence 521, App
31	1636	100.0	2974	9	US-10-184-658-521	Sequence 521, App
32	1636	100.0	2974	9	US-10-173-695-521	Sequence 521, App
33	1636	100.0	2974	9	US-10-173-697-521	Sequence 521, App
34	1636	100.0	2974	9	US-10-173-705-521	Sequence 521, App
35	1636	100.0	2974	9	US-10-174-576-521	Sequence 521, App
36	1636	100.0	2974	9	US-10-174-585-521	Sequence 521, App
37	1636	100.0	2974	9	US-10-174-586-521	Sequence 521, App
38	1636	100.0	2974	9	US-10-175-747-521	Sequence 521, App
39	1636	100.0	2974	9	US-10-176-481-521	Sequence 521, App
40	1636	100.0	2974	9	US-10-176-485-521	Sequence 521, App
41	1636	100.0	2974	9	US-10-176-487-521	Sequence 521, App
42	1636	100.0	2974	9	US-10-176-493-521	Sequence 521, App
43	1636	100.0	2974	9	US-10-176-756-521	Sequence 521, App
44	1636	100.0	2974	9	US-10-176-911-521	Sequence 521, App
45	1636	100.0	2974	9	US-10-176-919-521	Sequence 521, App

ALIGNMENTS

RESULT 1
US-09-981-353-165
; Sequence 165, Application US/09981353
; Patent No. US20020160382A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy W.
; APPLICANT: Jones, David A.
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
; FILE REFERENCE: PA-0038 US
; CURRENT APPLICATION NUMBER: US/09/981,353
; CURRENT FILING DATE: 2001-10-11
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 165
; LENGTH: 1636
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020160382A1 2434655CB1
US-09-981-353-165

Query Match 100.0%; Score 1636; DB 9; Length 1636;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1636; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GATCAGTGTGTGAGGGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GATCAGTGTGTGAGGGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGC 60

Qy         61 TCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTG 120

Qy        121 ACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATG 180

Qy        181 AGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCAT 240
```

```
Qy       241 TGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTG 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 TGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTG 300

Qy       301 ACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATG 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 ACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATG 360

Qy       361 ATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATC 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 ATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATC 420

Qy       421 AGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGA 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 AGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGA 480

Qy       481 TTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAA 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 TTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAA 540

Qy       541 TTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATG 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 TTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATG 600

Qy       601 TACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATT 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 TACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATT 660

Qy       661 CAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAG 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 CAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAG 720

Qy       721 AGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTG 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 AGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTG 780

Qy       781 AGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACT 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 AGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACT 840

Qy       841 TTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAA 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 TTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAA 900

Qy       901 ATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTC 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 ATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTC 960

Qy       961 AAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGA 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 AAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGA 1020

Qy      1021 AGGTGTTATGGAGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGT 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1021 AGGTGTTATGGAGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGT 1080

Qy      1081 ATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTC 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1081 ATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTC 1140

Qy      1141 ATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTC 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1141 ATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTC 1200

Qy      1201 CCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAG 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1201 CCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAG 1260

Qy      1261 AAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTA 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1261 AAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTA 1320

Qy      1321 CCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTG 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1321 CCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTG 1380

Qy      1381 TAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCA 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1381 TAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCA 1440

Qy      1441 AGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGA 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1441 AGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGA 1500

Qy      1501 TTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTAT 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1501 TTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTAT 1560

Qy      1561 TTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1561 TTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1620

Qy      1621 ATTCAAGAAAGACCTG 1636
            ||||||||||||||||
Db      1621 ATTCAAGAAAGACCTG 1636
```

RESULT 2
US-09-981-353-33
; Sequence 33, Application US/09981353
; Patent No. US20020160382A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy W.
; APPLICANT: Jones, David A.
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
; FILE REFERENCE: PA-0038 US
; CURRENT APPLICATION NUMBER: US/09/981,353
; CURRENT FILING DATE: 2001-10-11
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 33
; LENGTH: 2966
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020160382A1 997080.1
US-09-981-353-33

Query Match 100.0%; Score 1636; DB 9; Length 2966;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1636; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GATCAGTGTGTGAGGGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          4 GATCAGTGTGTGAGGGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGC 63

Qy         61 TCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         64 TCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTG 123

Qy        121 ACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        124 ACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATG 183

Qy        181 AGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        184 AGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCAT 243

Qy        241 TGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        244 TGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTG 303

Qy        301 ACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        304 ACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATG 363
```

```
Qy       361 ATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATC 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       364 ATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATC 423

Qy       421 AGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGA 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       424 AGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGA 483

Qy       481 TTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAA 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       484 TTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAA 543

Qy       541 TTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATG 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       544 TTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATG 603

Qy       601 TACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATT 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       604 TACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATT 663

Qy       661 CAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAG 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       664 CAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAG 723

Qy       721 AGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTG 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       724 AGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTG 783

Qy       781 AGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACT 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       784 AGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACT 843

Qy       841 TTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAA 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       844 TTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAA 903

Qy       901 ATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTC 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       904 ATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTC 963

Qy       961 AAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGA 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       964 AAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGA 1023

Qy      1021 AGGTGTTATGGAGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGT 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1024 AGGTGTTATGGAGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGT 1083

Qy      1081 ATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTC 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1084 ATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTC 1143

Qy      1141 ATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTC 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1144 ATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTC 1203

Qy      1201 CCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAG 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1204 CCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAG 1263

Qy      1261 AAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTA 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1264 AAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTA 1323

Qy      1321 CCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTG 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1324 CCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTG 1383

Qy      1381 TAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCA 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1384 TAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCA 1443

Qy      1441 AGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGA 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1444 AGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGA 1503

Qy      1501 TTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTAT 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1504 TTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTAT 1563

Qy      1561 TTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1564 TTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1623

Qy      1621 ATTCAAGAAAGACCTG 1636
            ||||||||||||||||
Db      1624 ATTCAAGAAAGACCTG 1639
```

RESULT 3
US-10-174-590-521
; Sequence 521, Application US/10174590
; Publication No. US20030008352A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C42
; CURRENT APPLICATION NUMBER: US/10/174,590
; CURRENT FILING DATE: 2002-06-18
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 521
; LENGTH: 2974
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-174-590-521

Query Match 100.0%; Score 1636; DB 9; Length 2974;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1636; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         1 GATCAGTGTGTGAGGGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGC 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 GATCAGTGTGTGAGGGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGC 60

Qy        61 TCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTG 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 TCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTG 120

Qy       121 ACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATG 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 ACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATG 180

Qy       181 AGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCAT 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 AGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCAT 240

Qy       241 TGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTG 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 TGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTG 300

Qy       301 ACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATG 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 ACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATG 360
```

```
Qy       361 ATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATC 420
Db       361 ATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATC 420

Qy       421 AGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGA 480
Db       421 AGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGA 480

Qy       481 TTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAA 540
Db       481 TTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAA 540

Qy       541 TTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATG 600
Db       541 TTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATG 600

Qy       601 TACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATT 660
Db       601 TACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATT 660

Qy       661 CAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAG 720
Db       661 CAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAG 720

Qy       721 AGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTG 780
Db       721 AGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTG 780

Qy       781 AGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACT 840
Db       781 AGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACT 840

Qy       841 TTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAA 900
Db       841 TTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAA 900

Qy       901 ATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTC 960
Db       901 ATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTC 960

Qy       961 AAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGA 1020
Db       961 AAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGA 1020

Qy      1021 AGGTGTTATGGAGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGT 1080
Db      1021 AGGTGTTATGGAGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGT 1080

Qy      1081 ATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTC 1140
Db      1081 ATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTC 1140

Qy      1141 ATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTC 1200
Db      1141 ATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTC 1200

Qy      1201 CCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAG 1260
Db      1201 CCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAG 1260

Qy      1261 AAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTA 1320
Db      1261 AAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTA 1320

Qy      1321 CCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTG 1380
Db      1321 CCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTG 1380

Qy      1381 TAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCA 1440
Db      1381 TAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCA 1440

Qy      1441 AGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGA 1500
Db      1441 AGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGA 1500

Qy      1501 TTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTAT 1560
Db      1501 TTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTAT 1560

Qy      1561 TTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1620
Db      1561 TTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1620

Qy      1621 ATTCAAGAAAGACCTG 1636
Db      1621 ATTCAAGAAAGACCTG 1636
```

RESULT 4
US-10-176-758-521
; Sequence 521, Application US/10176758
; Publication No. US20030008353A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C104
; CURRENT APPLICATION NUMBER: US/10/176,758
; CURRENT FILING DATE: 2002-06-21
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 521
; LENGTH: 2974
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-176-758-521

Query Match 100.0%; Score 1636; DB 9; Length 2974;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1636; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GATCAGTGTGTGAGGGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGC 60
Db          1 GATCAGTGTGTGAGGGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGC 60

Qy         61 TCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTG 120
Db         61 TCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTG 120

Qy        121 ACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATG 180
Db        121 ACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATG 180

Qy        181 AGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCAT 240
Db        181 AGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCAT 240

Qy        241 TGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTG 300
Db        241 TGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTG 300

Qy        301 ACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATG 360
Db        301 ACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATG 360
```

```
Qy       361 ATTTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATC 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 ATTTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATC 420

Qy       421 AGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGA 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 AGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGA 480

Qy       481 TTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAA 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 TTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAA 540

Qy       541 TTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATG 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 TTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATG 600

Qy       601 TACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATT 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 TACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATT 660

Qy       661 CAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAG 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 CAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAG 720

Qy       721 AGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTG 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 AGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTG 780

Qy       781 AGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACT 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 AGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACT 840

Qy       841 TTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAA 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 TTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAA 900

Qy       901 ATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTC 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 ATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTC 960

Qy       961 AAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGA 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 AAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGA 1020

Qy      1021 AGGTGTTATGGAGGTACAAAGGAAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGT 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1021 AGGTGTTATGGAGGTACAAAGGAAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGT 1080

Qy      1081 ATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTC 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1081 ATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTC 1140

Qy      1141 ATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTC 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1141 ATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTC 1200

Qy      1201 CCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAG 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1201 CCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAG 1260

Qy      1261 AAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTA 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1261 AAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTA 1320

Qy      1321 CCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTG 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1321 CCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTG 1380

Qy      1381 TAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCA 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1381 TAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCA 1440

Qy      1441 AGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGA 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1441 AGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGA 1500

Qy      1501 TTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTAT 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1501 TTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTAT 1560

Qy      1561 TTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1561 TTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1620

Qy      1621 ATTCAAGAAAGACCTG 1636
            ||||||||||||||||
Db      1621 ATTCAAGAAAGACCTG 1636
```

RESULT 5
US-10-175-737-521
; Sequence 521, Application US/10175737
; Publication No. US20030013153A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C50
; CURRENT APPLICATION NUMBER: US/10/175,737
; CURRENT FILING DATE: 2002-06-19
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 521
; LENGTH: 2974
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-175-737-521

Query Match 100.0%; Score 1636; DB 9; Length 2974;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1636; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         1 GATCAGTGTGTGAGGGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGC 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 GATCAGTGTGTGAGGGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGC 60

Qy        61 TCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTG 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 TCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTG 120

Qy       121 ACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATG 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 ACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATG 180

Qy       181 AGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCAT 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 AGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCAT 240

Qy       241 TGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTG 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 TGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTG 300

Qy       301 ACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATG 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 ACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATG 360
```

```
Qy       361 ATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATC 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 ATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATC 420

Qy       421 AGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGA 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 AGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGA 480

Qy       481 TTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAA 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 TTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAA 540

Qy       541 TTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATG 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 TTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATG 600

Qy       601 TACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATT 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 TACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATT 660

Qy       661 CAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAG 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 CAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAG 720

Qy       721 AGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTG 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 AGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTG 780

Qy       781 AGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACT 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 AGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACT 840

Qy       841 TTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAA 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 TTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAA 900

Qy       901 ATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTC 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 ATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTC 960

Qy       961 AAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGA 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 AAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGA 1020

Qy      1021 AGGTGTTATGGAGGTACAAAGGAAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGT 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1021 AGGTGTTATGGAGGTACAAAGGAAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGT 1080

Qy      1081 ATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTC 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1081 ATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTC 1140

Qy      1141 ATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTC 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1141 ATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTC 1200

Qy      1201 CCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAG 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1201 CCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAG 1260

Qy      1261 AAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTA 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1261 AAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTA 1320

Qy      1321 CCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTG 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1321 CCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTG 1380

Qy      1381 TAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCA 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1381 TAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCA 1440

Qy      1441 AGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGA 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1441 AGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGA 1500

Qy      1501 TTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTAT 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1501 TTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTAT 1560

Qy      1561 TTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1561 TTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1620

Qy      1621 ATTCAAGAAAGACCTG 1636
            ||||||||||||||||
Db      1621 ATTCAAGAAAGACCTG 1636
```

RESULT 6
US-10-173-706-521
; Sequence 521, Application US/10173706
; Publication No. US20030022293A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C7
; CURRENT APPLICATION NUMBER: US/10/173,706
; CURRENT FILING DATE: 2002-06-17
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 521
; LENGTH: 2974
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-173-706-521

Query Match 100.0%; Score 1636; DB 9; Length 2974;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1636; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GATCAGTGTGTGAGGGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GATCAGTGTGTGAGGGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGC 60

Qy         61 TCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTG 120

Qy        121 ACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATG 180

Qy        181 AGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCAT 240

Qy        241 TGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTG 300

Qy        301 ACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATG 360
```

```
Qy       361 ATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATC 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 ATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATC 420

Qy       421 AGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGA 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 AGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGA 480

Qy       481 TTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAA 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 TTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAA 540

Qy       541 TTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATG 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 TTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATG 600

Qy       601 TACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATT 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 TACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATT 660

Qy       661 CAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAG 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 CAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAG 720

Qy       721 AGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTG 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 AGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTG 780

Qy       781 AGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACT 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 AGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACT 840

Qy       841 TTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAA 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 TTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAA 900

Qy       901 ATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTC 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 ATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTC 960

Qy       961 AAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGA 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 AAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGA 1020

Qy      1021 AGGTGTTATGGAGGTACAAAGGAAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGT 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1021 AGGTGTTATGGAGGTACAAAGGAAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGT 1080

Qy      1081 ATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTC 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1081 ATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTC 1140

Qy      1141 ATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTC 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1141 ATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTC 1200

Qy      1201 CCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAG 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1201 CCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAG 1260

Qy      1261 AAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTA 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1261 AAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTA 1320

Qy      1321 CCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTG 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1321 CCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTG 1380

Qy      1381 TAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCA 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1381 TAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCA 1440

Qy      1441 AGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGA 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1441 AGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGA 1500

Qy      1501 TTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTAT 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1501 TTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTAT 1560

Qy      1561 TTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1561 TTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1620

Qy      1621 ATTCAAGAAAGACCTG 1636
            ||||||||||||||||
Db      1621 ATTCAAGAAAGACCTG 1636
```

RESULT 7
US-10-175-738-521
; Sequence 521, Application US/10175738
; Publication No. US20030022294A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C45
; CURRENT APPLICATION NUMBER: US/10/175,738
; CURRENT FILING DATE: 2002-06-19
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 521
; LENGTH: 2974
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-175-738-521

Query Match 100.0%; Score 1636; DB 9; Length 2974;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1636; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GATCAGTGTGTGAGGGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GATCAGTGTGTGAGGGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGC 60

Qy         61 TCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTG 120

Qy        121 ACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATG 180

Qy        181 AGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCAT 240

Qy        241 TGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTG 300

Qy        301 ACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATG 360
```

```
Qy        361 ATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATC 420

Qy        421 AGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGA 480

Qy        481 TTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAA 540

Qy        541 TTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATG 600

Qy        601 TACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATT 660

Qy        661 CAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAG 720

Qy        721 AGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTG 780

Qy        781 AGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACT 840

Qy        841 TTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAA 900

Qy        901 ATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTC 960

Qy        961 AAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 AAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGA 1020

Qy       1021 AGGTGTTATGGAGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGGTGTTATGGAGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGT 1080

Qy       1081 ATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTC 1140

Qy       1141 ATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTC 1200

Qy       1201 CCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAG 1260

Qy       1261 AAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTA 1320

Qy       1321 CCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTG 1380

Qy       1381 TAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCA 1440
```

```
Qy       1441 AGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGA 1500

Qy       1501 TTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTAT 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTAT 1560

Qy       1561 TTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1620

Qy       1621 ATTCAAGAAAGACCTG 1636
              ||||||||||||||||
Db       1621 ATTCAAGAAAGACCTG 1636
```

RESULT 8
US-10-175-752-521
; Sequence 521, Application US/10175752.
; Publication No. US20030022295A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C60
; CURRENT APPLICATION NUMBER: US/10/175,752
; CURRENT FILING DATE: 2002-06-19
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 521
; LENGTH: 2974
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-175-752-521

Query Match 100.0%; Score 1636; DB 9; Length 2974;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1636; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GATCAGTGTGTGAGGGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GATCAGTGTGTGAGGGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGC 60

Qy         61 TCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTG 120

Qy        121 ACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATG 180

Qy        181 AGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCAT 240

Qy        241 TGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTG 300

Qy        301 ACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATG 360
```

```
Qy       361 ATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATC 420
Db       361 ATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATC 420

Qy       421 AGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGA 480
Db       421 AGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGA 480

Qy       481 TTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAA 540
Db       481 TTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAA 540

Qy       541 TTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATG 600
Db       541 TTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATG 600

Qy       601 TACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATT 660
Db       601 TACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATT 660

Qy       661 CAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAG 720
Db       661 CAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAG 720

Qy       721 AGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTG 780
Db       721 AGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTG 780

Qy       781 AGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACT 840
Db       781 AGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACT 840

Qy       841 TTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAA 900
Db       841 TTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAA 900

Qy       901 ATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTC 960
Db       901 ATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTC 960

Qy       961 AAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGA 1020
Db       961 AAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGA 1020

Qy      1021 AGGTGTTATGGAGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGT 1080
Db      1021 AGGTGTTATGGAGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGT 1080

Qy      1081 ATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTC 1140
Db      1081 ATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTC 1140

Qy      1141 ATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTC 1200
Db      1141 ATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTC 1200

Qy      1201 CCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAG 1260
Db      1201 CCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAG 1260

Qy      1261 AAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTA 1320
Db      1261 AAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTA 1320

Qy      1321 CCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTG 1380
Db      1321 CCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTG 1380

Qy      1381 TAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCA 1440
Db      1381 TAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCA 1440

Qy      1441 AGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGA 1500
Db      1441 AGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGA 1500

Qy      1501 TTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTAT 1560
Db      1501 TTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTAT 1560

Qy      1561 TTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1620
Db      1561 TTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1620

Qy      1621 ATTCAAGAAAGACCTG 1636
Db      1621 ATTCAAGAAAGACCTG 1636
```

RESULT 9
US-10-176-482-521
; Sequence 521, Application US/10176482
; Publication No. US20030022296A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C70
; CURRENT APPLICATION NUMBER: US/10/176,482
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 521
; LENGTH: 2974
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-176-482-521

Query Match 100.0%; Score 1636; DB 9; Length 2974;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1636; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GATCAGTGTGTGAGGGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGC 60
Db          1 GATCAGTGTGTGAGGGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGC 60

Qy         61 TCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTG 120
Db         61 TCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTG 120

Qy        121 ACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATG 180
Db        121 ACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATG 180

Qy        181 AGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCAT 240
Db        181 AGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCAT 240

Qy        241 TGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTG 300
Db        241 TGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTG 300

Qy        301 ACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATG 360
Db        301 ACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATG 360
```

```
Qy       361 ATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATC 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 ATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATC 420

Qy       421 AGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGA 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 AGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGA 480

Qy       481 TTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAA 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 TTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAA 540

Qy       541 TTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATG 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 TTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATG 600

Qy       601 TACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATT 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 TACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATT 660

Qy       661 CAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAG 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 CAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAG 720

Qy       721 AGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTG 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 AGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTG 780

Qy       781 AGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACT 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 AGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACT 840

Qy       841 TTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAA 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 TTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAA 900

Qy       901 ATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTC 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 ATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTC 960

Qy       961 AAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGA 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 AAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGA 1020

Qy      1021 AGGTGTTATGGAGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGT 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1021 AGGTGTTATGGAGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGT 1080

Qy      1081 ATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTC 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1081 ATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTC 1140

Qy      1141 ATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTC 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1141 ATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTC 1200

Qy      1201 CCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAG 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1201 CCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAG 1260

Qy      1261 AAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTA 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1261 AAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTA 1320

Qy      1321 CCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTG 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1321 CCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTG 1380

Qy      1381 TAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCA 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1381 TAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCA 1440

Qy      1441 AGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGA 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1441 AGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGA 1500

Qy      1501 TTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTAT 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1501 TTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTAT 1560

Qy      1561 TTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1561 TTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1620

Qy      1621 ATTCAAGAAAGACCTG 1636
            ||||||||||||||||
Db      1621 ATTCAAGAAAGACCTG 1636
```

RESULT 10
US-10-176-757-521
; Sequence 521, Application US/10176757
; Publication No. US20030022297A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C86
; CURRENT APPLICATION NUMBER: US/10/176,757
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 521
; LENGTH: 2974
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-176-757-521

Query Match 100.0%; Score 1636; DB 9; Length 2974;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1636; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GATCAGTGTGTGAGGGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GATCAGTGTGTGAGGGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGC 60

Qy         61 TCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTG 120

Qy        121 ACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATG 180

Qy        181 AGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCAT 240

Qy        241 TGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTG 300

Qy        301 ACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATG 360
```

```
Qy       361 ATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATC 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 ATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATC 420

Qy       421 AGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGA 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 AGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGA 480

Qy       481 TTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAA 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 TTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAA 540

Qy       541 TTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATG 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 TTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATG 600

Qy       601 TACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATT 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 TACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATT 660

Qy       661 CAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAG 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 CAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAG 720

Qy       721 AGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTG 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 AGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTG 780

Qy       781 AGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACT 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 AGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACT 840

Qy       841 TTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAA 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 TTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAA 900

Qy       901 ATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTC 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 ATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTC 960

Qy       961 AAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGA 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 AAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGA 1020

Qy      1021 AGGTGTTATGGAGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGT 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1021 AGGTGTTATGGAGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGT 1080

Qy      1081 ATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTC 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1081 ATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTC 1140

Qy      1141 ATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTC 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1141 ATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTC 1200

Qy      1201 CCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAG 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1201 CCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAG 1260

Qy      1261 AAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTA 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1261 AAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTA 1320

Qy      1321 CCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTG 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1321 CCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTG 1380

Qy      1381 TAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCA 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1381 TAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCA 1440

Qy      1441 AGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGA 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1441 AGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGA 1500

Qy      1501 TTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTAT 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1501 TTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTAT 1560

Qy      1561 TTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1561 TTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1620

Qy      1621 ATTCAAGAAAGACCTG 1636
            ||||||||||||||||
Db      1621 ATTCAAGAAAGACCTG 1636
```

RESULT 11
US-10-176-913-521
; Sequence 521, Application US/10176913
; Publication No. US20030022298A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C66
; CURRENT APPLICATION NUMBER: US/10/176,913
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See file Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 521
; LENGTH: 2974
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-176-913-521

Query Match 100.0%; Score 1636; DB 9; Length 2974;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1636; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GATCAGTGTGTGAGGGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GATCAGTGTGTGAGGGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGC 60

Qy         61 TCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTG 120

Qy        121 ACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATG 180

Qy        181 AGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCAT 240

Qy        241 TGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTG 300

Qy        301 ACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATG 360
```

```
Qy       361 ATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATC 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 ATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATC 420

Qy       421 AGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGA 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 AGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGA 480

Qy       481 TTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAA 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 TTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAA 540

Qy       541 TTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATG 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 TTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATG 600

Qy       601 TACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATT 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 TACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATT 660

Qy       661 CAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAG 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 CAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAG 720

Qy       721 AGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTG 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 AGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTG 780

Qy       781 AGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACT 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 AGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACT 840

Qy       841 TTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAA 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 TTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAA 900

Qy       901 ATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTC 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 ATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTC 960

Qy       961 AAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGA 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 AAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGA 1020

Qy      1021 AGGTGTTATGGAGGTACAAAGGAAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGT 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1021 AGGTGTTATGGAGGTACAAAGGAAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGT 1080

Qy      1081 ATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTC 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1081 ATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTC 1140

Qy      1141 ATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTC 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1141 ATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTC 1200

Qy      1201 CCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAG 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1201 CCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAG 1260

Qy      1261 AAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTA 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1261 AAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTA 1320

Qy      1321 CCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTG 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1321 CCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTG 1380

Qy      1381 TAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCA 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1381 TAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCA 1440

Qy      1441 AGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGA 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1441 AGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGA 1500

Qy      1501 TTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTAT 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1501 TTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTAT 1560

Qy      1561 TTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1561 TTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1620

Qy      1621 ATTCAAGAAAGACCTG 1636
            ||||||||||||||||
Db      1621 ATTCAAGAAAGACCTG 1636
```

RESULT 12
US-10-180-552-521
; Sequence 521, Application US/10180552
; Publication No. US20030022300A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C153
; CURRENT APPLICATION NUMBER: US/10/180,552
; CURRENT FILING DATE: 2002-06-25
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 521
; LENGTH: 2974
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-180-552-521

Query Match 100.0%; Score 1636; DB 9; Length 2974;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1636; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GATCAGTGTGTGAGGGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GATCAGTGTGTGAGGGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGC 60

Qy         61 TCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTG 120

Qy        121 ACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATG 180

Qy        181 AGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCAT 240

Qy        241 TGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTG 300

Qy        301 ACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATG 360
```

```
Qy       361 ATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATC 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 ATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATC 420

Qy       421 AGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGA 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 AGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGA 480

Qy       481 TTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAA 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 TTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAA 540

Qy       541 TTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATG 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 TTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATG 600

Qy       601 TACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATT 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 TACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATT 660

Qy       661 CAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAG 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 CAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAG 720

Qy       721 AGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTG 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 AGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTG 780

Qy       781 AGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACT 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 AGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACT 840

Qy       841 TTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAA 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 TTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAA 900

Qy       901 ATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTC 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 ATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTC 960

Qy       961 AAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGA 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 AAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGA 1020

Qy      1021 AGGTGTTATGGAGGTACAAAGGAAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGT 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1021 AGGTGTTATGGAGGTACAAAGGAAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGT 1080

Qy      1081 ATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTC 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1081 ATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTC 1140

Qy      1141 ATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTC 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1141 ATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTC 1200

Qy      1201 CCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAG 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1201 CCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAG 1260

Qy      1261 AAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTA 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1261 AAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTA 1320

Qy      1321 CCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTG 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1321 CCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTG 1380

Qy      1381 TAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCA 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1381 TAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCA 1440

Qy      1441 AGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGA 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1441 AGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGA 1500

Qy      1501 TTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTAT 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1501 TTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTAT 1560

Qy      1561 TTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1561 TTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1620

Qy      1621 ATTCAAGAAAGACCTG 1636
            ||||||||||||||||
Db      1621 ATTCAAGAAAGACCTG 1636
```

RESULT 13
US-10-180-557-521
; Sequence 521, Application US/10180557
; Publication No. US20030022301A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C147
; CURRENT APPLICATION NUMBER: US/10/180,557
; CURRENT FILING DATE: 2002-06-25
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 521
; LENGTH: 2974
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-180-557-521

Query Match 100.0%; Score 1636; DB 9; Length 2974;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1636; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GATCAGTGTGTGAGGGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GATCAGTGTGTGAGGGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGC 60

Qy         61 TCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTG 120

Qy        121 ACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATG 180

Qy        181 AGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCAT 240

Qy        241 TGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTG 300

Qy        301 ACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATG 360
```

```
Qy       361 ATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATC 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 ATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATC 420

Qy       421 AGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGA 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 AGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGA 480

Qy       481 TTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAA 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 TTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAA 540

Qy       541 TTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATG 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 TTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATG 600

Qy       601 TACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATT 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 TACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATT 660

Qy       661 CAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAG 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 CAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAG 720

Qy       721 AGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTG 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 AGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTG 780

Qy       781 AGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACT 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 AGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACT 840

Qy       841 TTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAA 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 TTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAA 900

Qy       901 ATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTC 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 ATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTC 960

Qy       961 AAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGA 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 AAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGA 1020

Qy      1021 AGGTGTTATGGAGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGT 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1021 AGGTGTTATGGAGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGT 1080

Qy      1081 ATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTC 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1081 ATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTC 1140

Qy      1141 ATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTC 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1141 ATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTC 1200

Qy      1201 CCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAG 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1201 CCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAG 1260

Qy      1261 AAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTA 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1261 AAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTA 1320

Qy      1321 CCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTG 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1321 CCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTG 1380

Qy      1381 TAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCA 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1381 TAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCA 1440

Qy      1441 AGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGA 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1441 AGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGA 1500

Qy      1501 TTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTAT 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1501 TTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTAT 1560

Qy      1561 TTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1561 TTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1620

Qy      1621 ATTCAAGAAAGACCTG 1636
            ||||||||||||||||
Db      1621 ATTCAAGAAAGACCTG 1636
```

RESULT 14
US-10-173-700-521
; Sequence 521, Application US/10173700
; Publication No. US20030027262A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C14
; CURRENT APPLICATION NUMBER: US/10/173,700
; CURRENT FILING DATE: 2002-06-17
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 521
; LENGTH: 2974
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-173-700-521

Query Match 100.0%; Score 1636; DB 9; Length 2974;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1636; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GATCAGTGTGTGAGGGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GATCAGTGTGTGAGGGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGC 60

Qy         61 TCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTG 120

Qy        121 ACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATG 180

Qy        181 AGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCAT 240

Qy        241 TGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTG 300

Qy        301 ACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATG 360
```

```
Qy       361 ATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATC 420
Db       361 ATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATC 420

Qy       421 AGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGA 480
Db       421 AGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGA 480

Qy       481 TTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAA 540
Db       481 TTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAA 540

Qy       541 TTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATG 600
Db       541 TTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATG 600

Qy       601 TACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATT 660
Db       601 TACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATT 660

Qy       661 CAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAG 720
Db       661 CAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAG 720

Qy       721 AGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTG 780
Db       721 AGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTG 780

Qy       781 AGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACT 840
Db       781 AGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACT 840

Qy       841 TTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAA 900
Db       841 TTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAA 900

Qy       901 ATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTC 960
Db       901 ATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTC 960

Qy       961 AAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGA 1020
Db       961 AAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGA 1020

Qy      1021 AGGTGTTATGGAGGTACAAAGGAAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGT 1080
Db      1021 AGGTGTTATGGAGGTACAAAGGAAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGT 1080

Qy      1081 ATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTC 1140
Db      1081 ATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTC 1140

Qy      1141 ATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTC 1200
Db      1141 ATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTC 1200

Qy      1201 CCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAG 1260
Db      1201 CCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAG 1260

Qy      1261 AAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTA 1320
Db      1261 AAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTA 1320

Qy      1321 CCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTG 1380
Db      1321 CCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTG 1380

Qy      1381 TAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCA 1440
Db      1381 TAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCA 1440

Qy      1441 AGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGA 1500
Db      1441 AGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGA 1500

Qy      1501 TTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTAT 1560
Db      1501 TTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTAT 1560

Qy      1561 TTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1620
Db      1561 TTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1620

Qy      1621 ATTCAAGAAAGACCTG 1636
Db      1621 ATTCAAGAAAGACCTG 1636
```

RESULT 15
US-10-174-572-521
; Sequence 521, Application US/10174572
; Publication No. US20030027263A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C40
; CURRENT APPLICATION NUMBER: US/10/174,572
; CURRENT FILING DATE: 2002-06-18
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 521
; LENGTH: 2974
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-174-572-521

Query Match 100.0%; Score 1636; DB 9; Length 2974;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1636; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GATCAGTGTGTGAGGGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGC 60
Db          1 GATCAGTGTGTGAGGGAACTGCCATCATGAGGTCTGACAAGTCAGCTTTGGTATTTCTGC 60

Qy         61 TCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTG 120
Db         61 TCCTGCAGCTCTTCTGTGTTGGCTGTGGATTCTGTGGGAAAGTCCTGGTGTGGCCCTGTG 120

Qy        121 ACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATG 180
Db        121 ACATGAGCCATTGGCTTAATGTCAAGGTCATTCTAGAAGAGCTCATAGTGAGAGGCCATG 180

Qy        181 AGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCAT 240
Db        181 AGGTAACAGTATTGACTCACTCAAAGCCTTCGTTAATTGACTACAGGAAGCCTTCTGCAT 240

Qy        241 TGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTG 300
Db        241 TGAAATTTGAGGTGGTCCATATGCCACAGGACAGAACAGAAGAAAATGAAATATTTGTTG 300

Qy        301 ACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATG 360
Db        301 ACCTAGCTCTGAATGTCTTGCCAGGCTTATCAACCTGGCAATCAGTTATAAAATTAAATG 360
```

```
Qy       361 ATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATC 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 ATTTTTTTGTTGAAATAAGAGGAACTTTAAAAATGATGTGTGAGAGCTTTATCTACAATC 420

Qy       421 AGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGA 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 AGACGCTTATGAAGAAGCTACAGGAAACCAACTACGATGTAATGCTTATAGACCCTGTGA 480

Qy       481 TTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAA 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 TTCCCTGTGGAGACCTGATGGCTGAGTTGCTTGCAGTCCCTTTTGTGCTCACACTTAGAA 540

Qy       541 TTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATG 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 TTTCTGTAGGAGGCAATATGGAGCGAAGCTGTGGGAAACTTCCAGCTCCACTTTCCTATG 600

Qy       601 TACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATT 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 TACCTGTGCCTATGACAGGACTAACAGACAGAATGACCTTTCTGGAAAGAGTAAAAAATT 660

Qy       661 CAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAG 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 CAATGCTTTCAGTTTTGTTCCACTTCTGGATTCAGGATTACGACTATCATTTTTGGGAAG 720

Qy       721 AGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTG 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 AGTTTTATAGTAAGGCATTAGGAAGGCCCACTACATTATGTGAGACTGTGGGAAAAGCTG 780

Qy       781 AGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACT 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 AGATATGGCTAATACGAACATATTGGGATTTTGAATTTCCTCAACCATACCAACCTAACT 840

Qy       841 TTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAA 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 TTGAGTTTGTTGGAGGATTGCACTGTAAACCTGCCAAAGCTTTGCCTAAGGAAATGGAAA 900

Qy       901 ATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTC 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 ATTTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGTTTTCTCTGGGGTCACTGTTTC 960

Qy       961 AAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGA 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 AAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCCACAGA 1020

Qy      1021 AGGTGTTATGGAGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGT 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1021 AGGTGTTATGGAGGTACAAAGGAAAAAAACCATCCACATTAGGAGCCAATACTCGGCTGT 1080

Qy      1081 ATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTC 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1081 ATGATTGGATACCCCAGAATGATCTTCTTGGTCATCCCAAAACCAAAGCTTTTATCACTC 1140

Qy      1141 ATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTC 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1141 ATGGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGGTGGGAGTTC 1200

Qy      1201 CCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAG 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1201 CCATATTTGGTGATCAGCTTGATAACATAGCTCACATGAAGGCCAAAGGAGCAGCTGTAG 1260

Qy      1261 AAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTA 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1261 AAATAAACTTCAAAACTATGACAAGCGAAGATTTACTGAGGGCTTTGAGAACAGTCATTA 1320

Qy      1321 CCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTG 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1321 CCGATTCCTCTTATAAAGAGAATGCTATGAGATTATCAAGAATTCACCATGATCAACCTG 1380

Qy      1381 TAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCA 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1381 TAAAGCCCCTAGATCGAGCAGTCTTCTGGATCGAGTTTGTCATGCGCCACAAAGGAGCCA 1440

Qy      1441 AGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGA 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1441 AGCACCTGCGATCAGCTGCCCATGACCTCACCTGGTTCCAGCACTACTCTATAGATGTGA 1500

Qy      1501 TTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTAT 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1501 TTGGGTTCCTGCTGACCTGTGTGGCAACTGCTATATTCTTGTTCACAAAATGTTTTTTAT 1560

Qy      1561 TTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1561 TTTCCTGTCAAAAATTTAATAAAACTAGAAAGATAGAAAAGAGGGAATAGATCTTTCCAA 1620

Qy      1621 ATTCAAGAAAGACCTG 1636
            ||||||||||||||||
Db      1621 ATTCAAGAAAGACCTG 1636
```

Search completed: July 16, 2003, 18:04:24
Job time : 220.273 secs